

Helping the Man in the Middle: Assessing and Training  
Referee Performance.

Duncan Mascarenhas

PhD.

The University of Edinburgh

2005



## AUTHOR'S DECLARATION

I declare that this thesis is entirely my own work. It is being submitted for the degree of Doctor of Philosophy at the University of Edinburgh and has not been submitted for any other award.

NAME: Duncan R D Mascarenhas

SIGNED:

DATE: May 25<sup>th</sup> 2005

## Abstract

This thesis describes an applied programme of research with the English Rugby Football Union (RFU) national panel of referees. Referee performance can have a critical impact on the outcomes of games and as a consequence the future status of the clubs & players concerned. Surprisingly, given the importance of referees, access to scientific support for these individuals is almost non-existent. Furthermore, there are only a handful of empirical investigations that have explored refereeing performance and none that has attempted to *train* sports officials. Accordingly, this programme of research set out to understand, operationalise and train the key aspects of referee performance. Using a multi-modal approach the first investigation used referee performance profiling and content analyses of three sources of literature to establish the key areas of referee performance. The Cornerstones Performance Model of Refereeing emerged, overarched by the psychological characteristics of excellence; featuring knowledge and application of the law; contextual judgment; personality and management skills; and fitness and positioning. The model was subsequently adopted by the RFU to structure the applied support programme and guide the development and selection of the English RFU referees. A naturalistic approach was adopted, focussing primarily on the decision-making aspects of the performance model. A video based, law-application assessment tool revealed surprisingly low levels of accuracy amongst referees and their support groups. Accordingly, a training programme was designed to reinforce accurate and coherent interpretations. A group of national panel referees watched videotaped scenarios taken from premier league games, showing 5 sets of 5 tackles, in each case with an expert providing the interpretation of the correct decision. All referee groups improved their performance from pre to posttest, with the lowest ranked referees showing significant improvements. However, as the performance model presents, referee DM is influenced by many factors beyond a simple application of the law. Accordingly, the final investigation explored the factors that change the game context and how they influence rugby-union referees management of the game. Following the nominal group technique, two groups of referees listed contextual factors that they felt might affect their decisions during a game. Individual ratings of

this list with both groups revealed the most important factors to be the “temper of the game,” “the level of player respect/rapport,” “position on pitch,” “scoreline,” and the “time left in the game.” To verify these factors a think-aloud protocol was conducted with three international referees, assessing how they weigh their decisions based on the context and how this affects their management of the game. The results suggest that elite referees use “preventative refereeing” to help maintain the natural flow to a game. Finally, the implications of this research programme are discussed in the light of expediting the development of high performance referees in open team sports.



# LIST OF CONTENTS

Author's Declaration	ii
Abstract	iii
List of Contents	v
List of Figures	ix
List of Tables	ix
List of Abbreviations	x
Acknowledgements	xi

## Chapter 1. Introduction & Overview of Refereeing

1.1 Context of the Thesis	1
1.2 Synopsis of RFU Support Programme	1
1.3 The Challenges Facing Rugby Union Refereeing	2
1.4 Academic Research on Officiating	5
1.41 Matter of Fact Decisions	7
1.42 Matter of Opinion Decisions	8
1.5 Summary of Research Findings	9
1.6 Structure of the Thesis	10

## Chapter 2. Developing a Model for Referee Performance

2.1 Preamble	12
2.2 Introduction	12
2.3 The Cornerstones Performance Model: Evolution and Validation	15
2.31 Phase 1: Developing the Model	15
2.311 <i>Assessor Reports</i>	16
2.312 <i>RFU Training Literature</i>	16
2.313 <i>Performance Profiling</i>	17
2.314 <i>Empirical Research</i>	17
2.315 <i>Data Analysis</i>	18
2.4 Results and Discussion (Phase 1)	19
2.41 Assessor Reports	20
2.42 RFU Training Literature	21
2.43 Performance Profiles	22
2.44 Empirical Research	25
2.5 Phase 2 Validating the Model	28
2.6 Results and Discussion (Phase 2)	28
2.7 Phase 3 Using the Model and General Discussion	29
2.8 A Systematic Approach to Examining the Model	31

## Chapter 3. Reviewing the Decision Making Literature

3.1 Preamble	33
3.2 The Demands of Rugby Union Refereeing	33
3.3 Perception and Vision Lessons from Motor Control Literature	34

3.4 Lessons from Judgement Literature in Sports Officiating	37
3.5 In Search of Appropriate DM Theory – Classical Decision Making?	37
3.51 Classical Decision Making	37
3.52 Naturalistic Decision Making	40
3.6 The Theoretical Underpinnings of NDM – How Do The Experts Do It?	43
3.7 The Development of Shared Mental Model Theory	46
3.8 Naturalistic Approaches to Training Coherent DM	48
3.9 Video Scenarios to Develop DM Performance	50
3.10 Developing DM in Rugby Union Refereeing: The Starting Point	51

#### Chapter 4. He May Not Always Be Right, but He's Always the Referee: A Method to Assess the Accuracy and Coherence of Rugby Union Refereeing

4.1 Preamble	52
4.2 Introduction	52
4.21 Measuring DM Performance – Accuracy, Coherence and Shared Mental Models	53
4.3 Method	57
4.31 Participants	57
4.32 Instruments	57
4.33 Pilot Testing	59
4.34 Procedure	59
4.35 Data Analysis	61
4.4 Results	62
4.41 Accuracy, Agreement, Coherence and Confidence Levels for All Participants	62
4.42 Accuracy, Agreement, Coherence and Confidence Levels by Group	63
4.43 Areas of Concern in Refereeing the Tackle	66
4.44 The Fidelity of the Video Recordings and the Naturalistic Paradigm	68
4.5 Discussion & Conclusions	68
4.51 Analyses of all Participants	68
4.52 Efficacy of SMMs to Test Officials' DM	69
4.53 Analyses by Group	70
4.54 Applied Implications	72
4.55 A Naturalistic Approach to Referee Decision-Making	74

#### Chapter 5. Training Accuracy and Coherence in Law Application

5.1 Preamble	76
5.2 Introduction	76
5.21 Developing Decision Making: Applying NDM to Refereeing	77
5.22 Training Accuracy and Coherence with Scenarios	78
5.23 Purpose of the Study	79

5.3 Method	79
5.31 Participants	79
5.32 Developing the Assessment and Coherence Training Tapes	81
5.321 <i>Pre and Posttest Assessment Tapes.</i>	82
5.322 <i>The Coherence Training Tape.</i>	82
5.33 Instrumentation	83
5.34 Procedure	84
5.341 <i>Pre and Posttests.</i>	84
5.342 <i>Experimental Group.</i>	85
5.343 <i>Control Groups.</i>	86
5.35 Data Analysis	86
5.4 Results	87
5.41 Quantitative Results	87
5.42 Referee Qualitative Feedback	90
5.5 Discussion	91
5.51 Why Did the Lower Ranked Referees Improve?	93
5.52 Why Didn't the Higher Ranked Referees Improve?	93
5.53 Why Only 50% Accuracy?	95

## Chapter 6. Referee Decision-Making: It's Not What You Give It's the Way that you Give It!

6.1 Preamble	98
6.2 Introduction	98
6.3 Phase One Method	102
6.31 Participants	102
6.32 Procedures	103
6.33 Data Confirmation	104
6.4 Phase One Results	104
6.5 Phase One Discussion	107
6.6 Phase Two Method	108
6.61 Participants	108
6.62 Pilot Study	108
6.63 Procedure	110
6.64 Data Analysis	111
6.7 Phase Two Results	112
6.71 Mismatches in Severity of Illegality and Management	113
6.72 Increases in Hardness of Verbal and Non-Verbal Management	114
6.8 Phase Two Discussion	118
6.9 General Discussion and Conclusions	119
6.91 Issues of Trustworthiness, Reliability and Validity	120
6.92 Implications	121
6.10 Conclusions	123

## Chapter 7. Other Developments from RFU Support

7.1 Personality & Management Skills Training	125
7.2 Subsequent Law Application Interventions	126
7.3 Touch-Judge Coherence Training	127
7.4 Touch-Judge Communication Training	129
7.5 Organisational Issues	131
7.6 Ongoing Development of the Performance Model	132
7.7 Elite Referee Support	135

## Chapter 8. Summary, Conclusions and Future Recommendations

8.1 Introduction	137
8.2 Summary of Thesis	137
8.3 The RFU Legacy	139
8.4 Contributions to the body of knowledge	141
8.5 Future Directions	143

## References

### Appendix One The Cornerstones Model Focus Group Schedule

### Appendix Two Chapter Four Response Sheet

### Appendix Three Chapter Five Response Sheet

### Appendix Four Chapter Five Coherence Training Booklet

### Appendix Five Chapter Five Instructions to Participants

### Appendix Six Chapter Five Acceptance Sheet

### Appendix Seven Chapter Six Focus Group Schedule

### Appendix Eight Chapter Six Rating Scale & Response Sheet

### Appendix Nine Publications and Conference Presentations

## LIST OF FIGURES

Figure 2.1 - Assessor Reports	23
Figure 2.2 - Performance Profiles	24
Figure 2.3 - The Cornerstones Performance Model of Refereeing	27
Figure 6.1 - In-Game Contextual Factors Identified by Groups of High Potential and Full-Time Referees	105
Figure 7.1 - Conflict Management Grid	126
Figure 7.2 - Training Emphasis Model on the Robotic and Humanistic Skills of Refereeing	133
Figure 7.3 - Talent Development and Responses to Pressure in Refereeing Skills	134

## LIST OF TABLES

Table 4.1 - Responses of All Participants Expressed as a Percentage	64
Table 4.2 - Percentage of Correct Responses, Agreement and Confidence Scores by Group	65
Table 5.1 - Percentage Accuracy by Group	89
Table 6.1 - Referees "Contextual Factor" Mean Importance Ratings (out of 10)	106
Table 6.2 - "Illegality of Decision" Rating Sheet	110
Table 6.3 - Hardness of Referees' Responses (From 1-4) Comparing Kickable and Non-Kickable Tackles	118

## LIST OF ABBREVIATIONS

DM -	Decision Making
DMM -	Declarative Mental Model
FA -	Football Association
FIBA -	Fédération Internationale de Basketball
iRB -	international Rugby Board
NASO -	National Association of Sports Officials
NDM -	Naturalistic Decision Making
NGB -	National Governing Body
PMM -	Procedural Mental Model
R&D -	Research and Development
RFU -	Rugby Football Union
SMM -	Shared Mental Model
TMO -	Television Match Official

# ACKNOWLEDGEMENTS

Dedicated to the memory of Bob Morris

1952 - 2004

Many people have made this possible. First, I would like to thank my supervisor and mentor, Dave Collins, who was the innovator behind this project, and whose inimitable style of support and encouragement helped me to not only complete this thesis but also achieve many other personal milestones. I am grateful for the contribution made by my 2<sup>nd</sup> supervisor, Bob Morris, who sadly died before I was able to submit the thesis but who became very interested in my work, was a tremendous help with the writing and always fun to visit.

Thanks to Patrick Mortimer who began the project with me and developed my enthusiasm for decision-making research and became a fine adversary for academic banter.

I'd also like to thank Nick Bunting at the RFU Referees Centre of Excellence for having the foresight to back this project, and Colin High, Ed Morrison, Campo, Spreaders, and Whitey who allowed us to run with our ideas and whose choice of words regularly provided great entertainment.

I am grateful to all the postgraduate researchers in the PESLS department, especially Trevor Kerplunk, Amanda Martindale, Dr Thong Show and Chris Connaboy for their random chats and support throughout, and to Steve and Useless the cat for sharing their home with me so that I could finish writing up.

Finally, thanks to mum & dad, (Jo and Pete) for their financial and emotional support and for somehow finding the persistence to encourage me in my 16-year career as a student. May it continue...

# Chapter 1. Introduction & Overview of Refereeing

## 1.1 Context of the Thesis

This thesis describes a scientifically validated programme of research with the RFU referees. As with any programme for elite performers, before beginning applied support it was necessary to complete a coherent analysis of the demands of that sport (Bull, 1989) to make assessments and develop interventions that will have a direct and positive impact upon performance. The aims were first to *understand* the performance environment by identifying the key areas of refereeing performance; second to *operationalise* a method for studying each area; and third develop interventions to *train* that area.

## 1.2 Synopsis of the RFU Support Programme

In July 1998 a team of five sport scientists, including myself,<sup>1</sup> began an empirical research driven support programme with England's Rugby Football Union (RFU) referee Centre of Excellence, initially focussing support on the top 65 referees, known as the national panel. This group of part-time officials were responsible for refereeing the premiership and four national leagues, together with a support group of touch-judges, advisers (responsible for assessing referee performance), coaches

---

<sup>1</sup> As this was a team project, in order to be clear about which aspects of work I completed and in conjunction with APA guidelines reflecting increased tolerance for personal pronouns (American Psychological Association, 2003, p. 39), the first person tense is occasionally used in this thesis.



(responsible for individual referee progress), and development officers. The programme of support and development was tapered so that the top-20 referees, largely operating at the premiership level, received the most support, although a key target for the sport science team was to make an impact throughout rugby union refereeing in England. A key element to the programme of support was a research and development (R&D) approach, and the need to produce practical training tools for the top referees that could be cascaded down throughout the RFU referees to make an impact throughout the organisation.

The sports science team adopted a “hands-on” approach, exemplified by regular attendance at both referee group and management group meetings. Developing this close working relationship, and building trust, rapport and credibility through numerous presentations at regional meetings, a flexible work programme was developed to adapt to arising demands on any of the individuals involved in the development of top-level rugby union officiating in England.

### 1.3 The Challenges Facing Rugby Union Refereeing

Refereeing team sports is becoming increasingly challenging, as more teams turn professional, media scrutiny increases and the game evolves rapidly. Unfortunately, in comparison to the players that they officiate, the referees appointed to adjudicate these games still receive comparatively little remuneration (Poll, 2003), sport science support (Ford, Gallagher, Lacy, Bridwell, & Goodwin, 1999), or appropriate training (see Garcia, 2003; Mascarenhas, MacPherson, Ollis & Collins, 2003; cf. Ste-Marie, 2003). To compound the situation, research conducted on such officials has rarely

considered factors that directly contribute to referee performance *per se*; accordingly, the implications for the advancement of refereeing are limited.

Furthermore, with the international Rugby Board's (iRB) removal of the declaration of amateurism from the laws<sup>2</sup> of the game in 1995 (iRB, 1995) England's top rugby union division, the premier league, became professional. Consequently, all the premier league clubs now have full-time managers, players, physiotherapists, and fitness coaches. Able to dedicate much more time to training, they are in much better physical condition (Eaves & Hughes, 2003) and their performance has soared, which may in part explain England's recent world cup success. As one consequence of such improved conditioning, the game has become faster and the players more cynical with regard the laws of the game devious (Ackford, 2002; Morrison, 2002). Despite all these player and team advancements, however, it was not until August 1999 that England employed their first full-time referees. Perpetuating this problem, only 5 of the 13 referees who officiated at premiership level in 2004 were full-time professionals. Addressing this possible mismatch in skill and professionalism is crucial, since a team's relegation or promotion has severe financial implications and decisions that referees make can decide the course of a player's career (Craven, 1998).

Parallel pressures arise from escalating media attention. As television production has developed, referees and touch-judges (their assistants who officiate the sidelines) have become increasingly aware of the camera angles that are used to scrutinise their

---

<sup>2</sup> Throughout this thesis the term laws is used synonymously with rules.

performance (C. White, personal communication, March 4, 2003). In order to adapt to the increased television presence, in 2001 the iRB introduced new laws to allow a television match official (TMO) to help the “on-field referee” by using slow-motion video replays to assist with uncertain decisions (iRB, 2003). The TMO typically sits in a video-production booth and reviews a range of angles, often identical to those broadcast to the general public, in order to reach a decision. This video-review process takes time to administer however, and although one might assume that it would take the pressure off the referees, it appears to have *contributed* to that pressure, as throughout the world cup they were criticised for taking too long to make a decision (Barwick, 2003).

Pressure on referees to perform also emerges from the national press and from their own referee national governing bodies (NGBs). Referees who do not perform well are under such scrutiny that they risk public humiliation in the daily newspapers or demotion by their NGB (The Guardian, 2001; Wilson, 2000). In addition, these same NGBs do not appear to have a clear focus on the key performance indicators for referees (see Sabatini, 2002). Commonly, across most major team sports, fitness testing through generic field tests (e.g., Ramsbottom, Brewer & Williams, 1988) is often the sole measure of aptitude, beyond occasional and infrequently conducted match assessments. These are usually conducted by senior ex-officials who become almost “guru-like” as there is no systematic procedure to their assessments and they frequently rely on gut feelings and intuition (Sabatini, 2003). When these assessments are more systematic, they often become overly structured into a long list of competencies (e.g., Griffiths, Dickinson & Renton, 1999) such that it diminishes

the task into a skeleton of itself (Lyle, 1999). Furthermore, when referees come to upgrade their level of certification, their knowledge and understanding of the law is invariably examined through pencil and paper tests (see Melrose, 2003). Typically, yearly conferences are the only time when more ecologically valid, video-based evidence is reviewed and the *application* of those laws is considered; even then however, there is seldom any systematic input to either development or evaluation. In short, at the time of the project's commencement the practical knowledge and techniques available for the evaluation and enhancement of referee performance was sadly lacking.

## 1.4 Academic Research on Officiating

To compound the challenge faced by those NGBs sufficiently enlightened to seek better methods of evaluation and enhancement, there is very little empirical research on refereeing. The few studies reported have typically surveyed perceived levels of stress and anxiety that referees experience in a variety of domains, including soccer (e.g., Taylor, Daniel, Leith & Burke, 1990), gridiron (American) football and volleyball (e.g., Goldsmith & Williams, 1992; e.g., Stewart & Ellery, 1996), baseball and softball (e.g., Rainey, 1995), and basketball in both the USA (e.g., Burke, Joyner, Pim & Czech, 2000; e.g., Rainey & Winterich, 1995) and Australia (e.g., Anshel & Weinberg, 1995; e.g., Kaissidis-Rodafinos, Anshel & Sideridis, 1998). Unfortunately, for the application of this knowledge to meet the purpose of performance enhancement, none of these papers has even attempted to assess how levels of stress may, or may not *directly affect* the performance of these officials. Indeed, it may be that stress is functional, allowing the referee to get activated for

good performance (Jokela & Hanin, 1999). In fact, the overwhelming majority of these studies reveal only very low to moderate levels of stress. It is remarkable however, that despite finding only very low levels of stress in baseball umpires and without an assessment of the impact of such stress, Rainey (1995) proposes a series of psychological interventions in order to reduce stress in sports officials (cf., Kerr, 1999) as a means of performance enhancement.

Similar themes of research have surveyed assaults on basketball referees (e.g., Rainey & Duggan, 1998) and rugby union referees (e.g., Rainey & Hardy, 1999a.). Again, although worthy of note, these studies have also failed to discuss the effect of such assaults upon performance, recruitment, retention or dropout of referees. Interestingly, when Rainey and Hardy (1999b) did assess the effect of stress on dropout in rugby union referees, they found that age and burnout were only marginally related to the intention to terminate and this relationship was merely correlative with no evidence of cause and effect.

In similar fashion, Kaissidis-Rodafinos, Anshel and Porter (1997) looked at how an inappropriate coping style might affect subsequent stressors; but once again (and as they concede) they did not directly consider how coping might affect performance. Thus, what literature there is available provides little assistance to NGBs concerned with recruiting, developing and retaining referees.

As a relatively recent development a few studies *have* looked directly at referee performance. These have investigated two types of situations, (1) matter of fact decisions and (2) matter of opinion decisions.

#### *1.41 Matter of Fact Decisions*

These are relatively closed, often “yes-or-no” types of decisions, determining between one outcome and another. For example, Craven (1999) commendably studied “leg-before-wicket” decisions in cricket, asking if the ball would hit the stumps or not. Craven discovered that feedback on performance could reduce the amount of errors made. Unfortunately, the use of only two subjects, both unqualified officials and one of whom had never umpired a cricket match before, makes it difficult to generalise the findings to expert officials.

Exploring a similar, “either-or” decision in baseball, Rainey, Larsen, Stephenson and Olson (1993) found evidence supporting the “phantom-tag” ruling, where umpires judging if a runner made it to the base before the ball, tended to call base runners out when the fielder had not tagged the base. In similar fashion, Oudejans, Verheijen, Gerrits, Steinbrückner and Beek (2000) found a parallax error in judging offside decisions in soccer, revealing important implications for officials’ positioning. Unfortunately, however, whilst these studies provide good foundations upon which to conduct investigations into matter-of-fact decisions in refereeing, the majority of decisions that have to be made in team sports are considerably more complex than this.

### 1.42 Matter of Opinion Decisions

Recently researchers have investigated more variable decisions that are not necessarily based on facts but rather which require an interpretation of events that have occurred. For example Nevill, Balmer and Williams (2002) presented tackle/challenge situations in soccer to 40 qualified referees, asking them to submit 1 of 4 options, either (1) home foul, (2) away foul, (3) no foul, or (4) uncertain. Nevill et al., found that those referees who watched the incidents in the presence of crowd noise were more likely to award fewer decisions against the home team, versus referees who watched without sound. Despite such valuable findings, due to their failure to report the level at which these referees officiated, it is impossible to contextualise the results; specifically, to know if the best referees would be equally influenced.

In similar fashion, Plessner & Betsch (2001) assessed the contingency effects of awarding penalties in soccer, after previous, similar “close-calls”. Unfortunately, without exposing the *reasons* for the increased likelihood of awarding a penalty after failing to grant a previous appeal, it is impossible to know if this occurred as a result of a contingency effect or some other more game-management focused rationale (Mascarenhas, Collins & Mortimer, 2002a).

The most laudable paper to date that assesses referee decision-making (DM), explored the impact that prior knowledge of a team’s aggressive reputation may have on a series of tackle/challenge situations in soccer (Jones, Paull and Erskine, 2002). They found that referees who were informed of a team’s aggressive reputation were

more likely to award red and yellow cards against that team than referees who did not receive such information. However, by failing to account for potential home advantage effects (see Nevill et al., 2002) or, more crucially, enquiring into the reasons underpinning such yellow and red cards, these findings may once again be somewhat distorted. Moreover, as in all these “matter-of-opinion” studies, the sample is reported as “qualified officials” rather than a clear exposition of expert status. Accordingly, the findings are difficult to generalise to the professional game which represents the primary focus of most NGBs.

A final issue with such studies is the relatively closed and constrained situations which have been examined. In open sports such as rugby union, many of the decisions that referees have to adjudicate are much more complex (Ackford, 2003) with considerably more degrees of freedom than the simple three or four option decisions used in the studies reported earlier. In their normal, ecologically valid situation, referees are challenged with multiple, complex and overlapping decisions with events occurring not only very quickly, but also at different places on the field, and often involving more than two players (Ackford, 2003). As such, in order to make a genuine contribution to performance, investigation must focus on more subtle and complex issues, such as those that typify the refereeing experience.

## 1.5 Summary of Research Findings

To date, there is no empirical research that has provided a method to help train the performance of expert referees; moreover, there is no clarity on what actually constitutes expert refereeing performance (see Stern, 2002). Anshel (1995) laudably



attempted to behaviourally quantify referee performance and made some valuable findings. However, in determining only *competence* in his refereeing rating-scale he may not have accounted for many of the attributes that lead to expertise.

Given such poor understanding of expert referee performance, it may be that Time Magazine's suggestion is true, proposing, "ideally the umpires should combine the integrity of a supreme court justice, the physical agility of an acrobat, the endurance of Job and the imperturbability of Buddha" (Time Magazine, 1961, p. 14). So, the main aim of the programme of investigation reported in this thesis was to address these various shortcomings and to enable a genuine contribution to demonstrable features of elite referee performance.

## 1.6 Structure of the Thesis

Detailing the 5 year support programme with the RFU referees, this thesis contains nine chapters. Just like many other referee NGBs, the only measure of performance for the RFU referees was the multistage fitness test (used to assess  $\text{VO}_2 \text{ max}$ ), although the recommended levels of achievement on this test were not scientifically validated. Given this lack of clarity, Chapter Two develops a performance model for refereeing and suggests interventions to train each of the key areas of the model. As the DM component of refereeing is crucial, Chapter Three explores approaches to DM research to find an appropriate paradigm to explore these aspects of the performance model. Consequently, adopting a naturalistic decision making (NDM) approach, Chapter Four explores arguably the most important cornerstone, law application. This phase develops a method to assess RFU officials' application of the

law to establish the levels of accuracy and coherence and highlight any potential problems. To improve performance in this area Chapter Five presents a referee coherence training programme through exposing an expert's knowledge and training a hierarchical system to referee DM.

Recognising the need to explore the “humanistic” elements of the model, Chapter Six investigates other factors that may affect referees' decisions using focus group interviews with a group of both high potential and international referees. Building upon these findings, a think-aloud protocol with the three highest ranked referees explores how they use their management skills to referee high-pressure games. Reflecting the need to be responsive to changing demands and produce scientifically validated tools to train every corner of the model Chapter Seven explains the revision of the work programme with the elite group and describes some of the other aspects of the work programme that emerged. Finally, a summary of each chapter is presented in Chapter Nine, focusing on the effects of the support programme within the RFU, the theoretical understanding that has emerged and the implications for other NGBs and recommendations for future research.

# Chapter 2. Developing a Model for Referee Performance

## 2.1 Preamble

In order to develop a coherent approach to support work, a model for support was required, focusing on areas that clearly and justifiably embrace referee performance. This would provide an operating language for discussing performance, a framework for selection, a structure for scientifically validated assessments and interventions, and the opportunity to develop coherence throughout the RFU with a clearer understanding of referee performance. In addition, while sport science research literature holds many investigations into the enhancement of elite athletes (e.g., Farrow & Abernethy, 2002), as suggested in the previous chapter there is a dearth of literature investigating the performance of match officials. As such, it was anticipated that this investigation would make a significant contribution to referee performance research.

## 2.2 Introduction

As noted earlier, the majority of studies into sports officials have considered avenues of psychological research, such as stress, coping, judgment bias, and personality characteristics of referees, which could be described as “researching psychology through refereeing”. While this might provide valuable insights into psychological phenomena, it does less to directly facilitate the performance of such sports officials; indeed this evolution seems to parallel much of the early work in sport psychology

where research used sport to address pertinent issues, rather than providing a primary focus on performance enhancement *per se*. Even in apparently specific training texts on officiating (e.g., Weinberg & Richardson, 1990) advice is often somewhat generic, reflecting ideas that are appropriate for all performers but lacking detailed attention to the essential and determining components of sports officiating. Moreover the literature that has looked directly at aspects of referee performance (e.g., Jones et al., 2002; Oudejans et al., 2000; Plessner & Betsch, 2001) has not investigated *expert* referees.

Given such a dearth of appropriate literature, it is necessary to consider how the RFU currently assesses and develops their referees. At the commencement of this project, in order to be promoted to the RFU refereeing national panel (top 65), referees would have to pass a written law exam and be put forward by their local refereeing society to a regional panel, offering them the chance to officiate the best games in their region. Based on their performance at this level, and subsequent to their successful completion of an RFU induction-day, they could then be invited to join the national panel. RFU assessors would measure the quality of their refereeing performance at these levels and complete a written report on each game. These reports were unstructured, providing a descriptive account of the game with occasional recommendations for future development (R. Debney, Top 5 national ranked and full-time referee, personal communication, September 21, 2005).

Similarly, when on the national panel, RFU assessors would provide written feedback to referees. Such unstructured continual assessments were the primary form of feedback they received, until the RFU adopted the “Competence Based Referee Development System” (Griffiths et al., 1999). This “tick-box” based system whilst an intuitively appealing way of assessing performance, was only capable of measuring *competency* in refereeing (as critiqued in Chapter One) and could not delineate between the elite and super-elite referees on the national panel (cf. Lyle).

These referees received no form of communication skills training, no specific law-based training, and no fitness training. In fact, there were only two requirements of these referees off the field of play, (1) to attend regional monthly meetings where referees discussed recent incidents that took place, sometimes with the assistance of videotape review, though these discussions were not structured, and (2) to perform a fitness test tri-annually, where referees had to reach level 11.5 on the “Multistage Fitness Test” (see Ramsbottom et al., 1988) in order to maintain their status as a national panel referee. Although attempts to ensure that referees maintain an *appropriate* level of fitness are to be commended, unfortunately, 11.5 appeared to be no more than an arbitrary level as this was not validated to the actual levels of fitness required to officiate rugby union (A. Melrose, RFU Rugby Referee Development Officer, personal communication, September 22, 2005).

Given that the expertise literature suggests that high levels of performance are reached

through deliberate practice totalling approximately 20 hours per week (Starkes, 2000), the absence of such training may lead us to question whether expert referees are being developed at all. Furthermore, since scientific support incorporating a range of physiological, biomechanical and psychological measures is now commonplace for elite athletes, it seems reasonable to provide similar support to referees. However, it is very difficult to support performance without a clear and justifiable understanding of its key components and contributory factors (see Taylor, 1995). In fact, Hardy & Parfitt (1994) suggest that an important feature of any sport psychology support is to identify the important psychological skills and teach them to the performer. Accordingly, as the first phase of a support program for the RFU referee department, the aim was to develop a referee performance model as an essential part of the needs analysis, so that assessment and training programs could be developed in areas that directly contribute to performance on the field of play. As a direct consequence, referees, and those responsible for training them would have a clear and empirically supported focus for deliberate practice interventions to improve refereeing performance (Ericsson, 2003).

## 2.3 The Cornerstones Performance Model:

### Evolution and Validation

#### *2.31 Phase 1: Developing the Model*

To accurately expose the components of referee performance in rugby union,

thematic content analyses were conducted on four sets of literature. The aim was to get systematic coverage of contrasting sources, triangulating not only data sources but also different groups of officials involved in the process to get a comprehensive picture of referee performance. These data sources were:

1. A random selection of 20 assessor reports written for 10 English national panel referees.
2. All the RFU training literature presented to the English RFU national panel of referees, touch-judges, assessors and referee coaches from January 1997 to December 2000 ( $n = 23$ ).
3. Performance profiling of a group of top-20 nationally ranked referees ( $n = 14$ ).
4. Published research articles taken from sport science journals ( $n = 58$ ).

*2.311 Assessor Reports.* Assessor reports, particularly for these elite referees often focus on hard to quantify but very important factors that seem to make the difference between good and elite referee performance. The two most recent assessor reports were taken from 10 randomly selected referees from the RFU national panel of referees, ensuring that at least two referees were taken from each level of the game, from premier league, down to national league division three.

*2.312 RFU Training Literature.* The RFU training literature and the assessment

system, known as the Competence Based Referee Development System (Griffiths et al., 1999) were content analysed in the same fashion as the assessor reports, seeking cogent and consistent themes. Since training materials need to be clear and understandable, these publications exposed these more formal aspects of referee performance, offering a useful contrast to the first category of data, which is typically presented in a more relaxed, conversational style.

*2.313 Performance Profiling.* 14 of the top-20 RFU panel referees took part in performance profiling at the English RFU referee national conference. The remaining six referees from the top-20 who did not take part were either on holiday or on international duty. Strictly following the guidelines prescribed by Butler and Hardy (1992), the concept was introduced to the referees, who then formed small groups of no more than five to brainstorm and discuss ideas and then share them with the whole squad. I then recorded a master list of constructs for the referees to rate their own performance, based on the group's suggestions and subsequent discussions. I was careful not to get in the way of participants expressing their opinions, experiences and suggestions, and encouraged all members to contribute. All the referees agreed that the final performance profile containing 17 characteristics represented an accurate, valid and valuable breakdown of refereeing skills, and many of the referees have since used the same list for personal development.

*2.314 Empirical Research.* During a period of over 6 months, a series of internet



and Sports Discuss searches were conducted to locate empirical research articles focusing on refereeing, judging, umpiring and sports officiating. In addition, the contents pages of The Sport Psychologist, the Journal of Sport and Exercise Psychology, the Journal of Applied Sport Psychology, the International Journal of Sport and Exercise Psychology, the Journal of Sport Sciences, and the Journal of Sport Behavior were inspected as these revealed the most articles following the initial searches. Each article that was found was checked for additional citations on sports officiating, revealing a total of 58 articles published between 1985 and 2002. However, in keeping with the arguments presented earlier, these data were considered critically against the justification presented for the salience of each topic to in-game performance. For example, research was only considered where either that paper or another empirical source could be found to justify the contribution, which the factor(s) under consideration made *directly* to officiating performance.

**2.315 Data Analysis.** The assessor reports, RFU training literature and performance profiles were analysed through a grounded theory approach (Strauss & Corbin, 1998). Following Hayes' (2000) guidelines on inductive thematic analysis, I read and reread all the reports, noting items of interest. All items that appeared to be dealing with similar topics were grouped together. I then systematically reread all the transcripts to ensure that themes or data had not been overlooked. When new potential themes emerged, I searched back through the data to check if other phrases under the new theme had been missed. To further validate this procedure, a second

researcher read the reports several times, and examined the allocation of themes. As described by Marshall and Rossman (1995) the second researcher served as a “devil’s advocate” challenging my perceptions and any potential bias that may have been drawn in. Systematically trawling through the data we discussed the meaning of phrases used by the assessors and subsequently examined each allocation to enhance reliability, trustworthiness and credibility. Based on the collective findings from the content analyses each theme was operationally defined by myself and subsequently verified by the second researcher and the sport psychology research team, paying particular attention to the choice of words used in each description.

## 2.4 Results and Discussion (Phase 1)

The inductive analysis of these four sources of data resulted in the emergence of five themes of refereeing performance. Four of the areas were largely psychologically based, and the fifth had more physiological underpinnings. The emergent themes were:

1. Knowledge and Application of the Law: The underpinning knowledge of the law that allows referees to accurately interpret dynamic situations and penalise accordingly.
2. Contextual Judgment: An appreciation of the tenor of the game and the referee’s ability to alter his or her style of refereeing to suit the particular nuances of the game.

3. Personality and Game Management: The verbal and non-verbal interpersonal skills that allow referees to communicate appropriately with the players.
4. Physical Fitness, and Positioning: The physical attributes that allow referees to “stay-up-with-play,” see the game from an appropriate perspective and use the appropriate signals.
5. Psychological Characteristics of Excellence: The common elements of successful performance (see McCaffrey & Orlick, 1989; Orlick & Partington, 1988) including commitment, goal setting, imagery, planning, distraction control, responses to pressure situations and realistic performance evaluations.

The assessor reports, RFU literature and performance profiles all consistently revealed these five cornerstone areas as crucial skills for referees. To date, there is only one other empirical research paper that has attempted to define refereeing performance. Anshel's (1995) “Behaviourally Anchored Rating Scale” of referee competencies exposed 13 characteristics as important for effective refereeing, 11 of which could distinguish between novice and expert referees. All 13 of these competencies are subsumed in the 5 performance categories identified in these results.

#### *2.41 Assessor Reports*

The assessor reports presented the most widespread information, with approximately 90% of the raw data falling under all the themes identified. The remaining 10% showed no coherence and due to lack of validation were not explored

further. Figure 2.1 includes three exemplars (per theme) of typical comments from the assessor forms. Without exception, each assessor's report made reference to psychological characteristics such as confidence, concentration, motivation, coping with pressure, and dealing with mistakes.

#### *2.42 RFU Training Literature*

Similar to the assessor report data the RFU literature was replete with very specific details on law interpretations. For example, yearly conference packs given to all the national panel officials (referees, touch-judges, assessors, and coaches) had clear guidelines under specific parts of the game, (viz., tackle, ruck/maul, scrummage, lineout, advantage, kick-off and restarts, offside, and the sin bin). In addition to these instructions on how to interpret and apply the law, officials were also guided on positioning skills and training regimes to develop their fitness. Five of these reports (22% of the sample) gave a modicum of advice on how to manage and communicate with players and coaches and one suggested the need for more communication between referees and coaches (Melrose, 1998). Another conference pack had a section entitled the "X-factor of refereeing," which described a performance factor that (apparently) eluded definition but which many super elite referees exhibit (Melrose, 1999). This appeared to embrace aspects of both game management and contextual judgment. For example, it was believed to be evident when "a referee has positive body language, especially at critical phases of the match" (p. 2) and also when top referees show "...the ability to recognise patterns of play and find

appropriate solutions, whilst showing empathy for the players and the context of the game” (p. 4). This factor (numerically small but subsequently recognised as crucial by all participants – see phase 2 of this chapter) highlighted the importance of recognising the interactions between the key areas, as well as discrete analysis, akin to observations of Bloom’s (1985) model for developing expertise.

### *2.43 Performance Profiles*

The performance profiling highlighted 17 characteristics of elite refereeing grounded within the referees’ understanding. These seemed to emphasise the same five areas that arose from the analysis of assessor reports and once again reinforced the need for an emphasis on psychological characteristics (see Figure 2.2).

*Figure 2.1 - Assessor Reports*

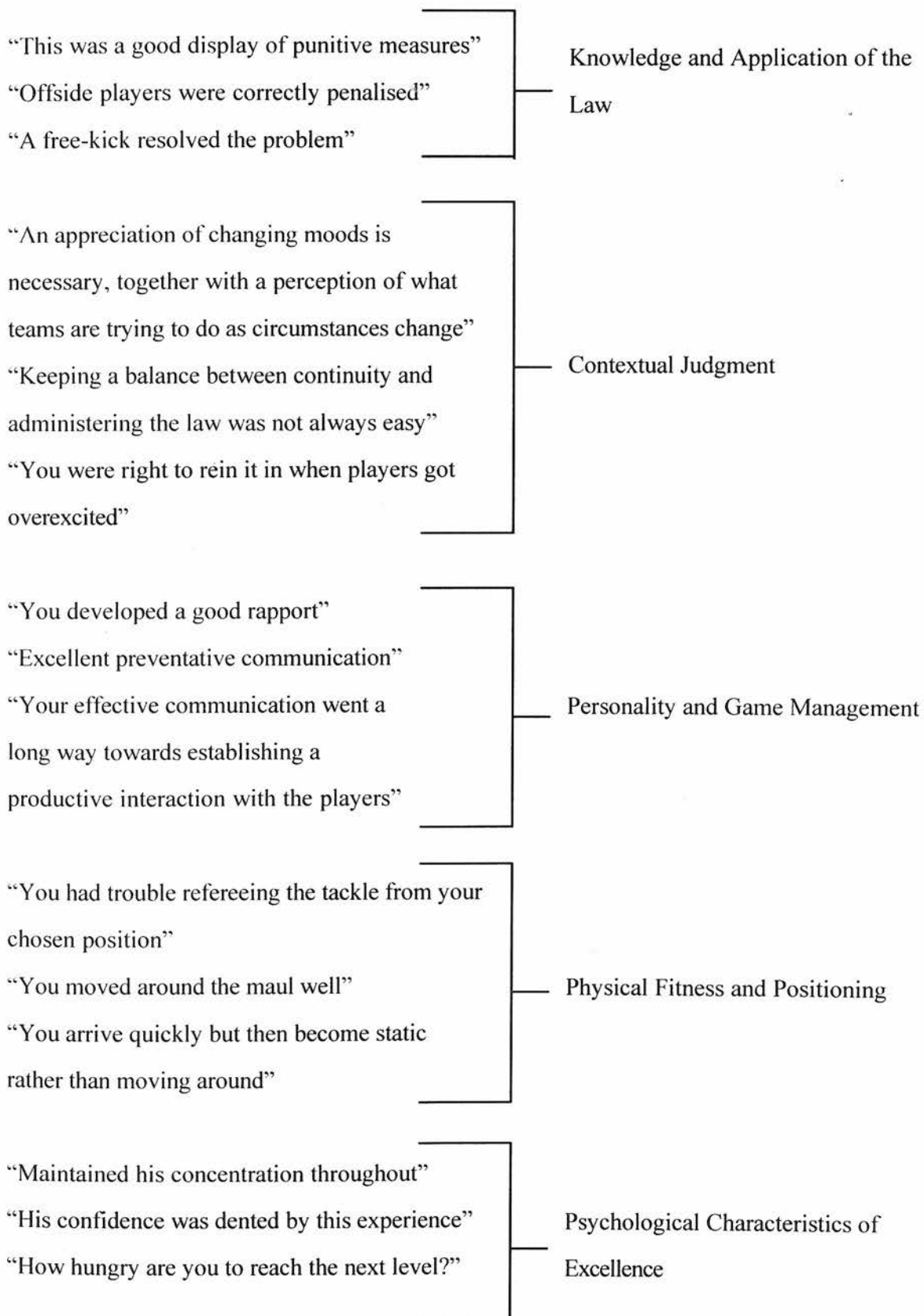
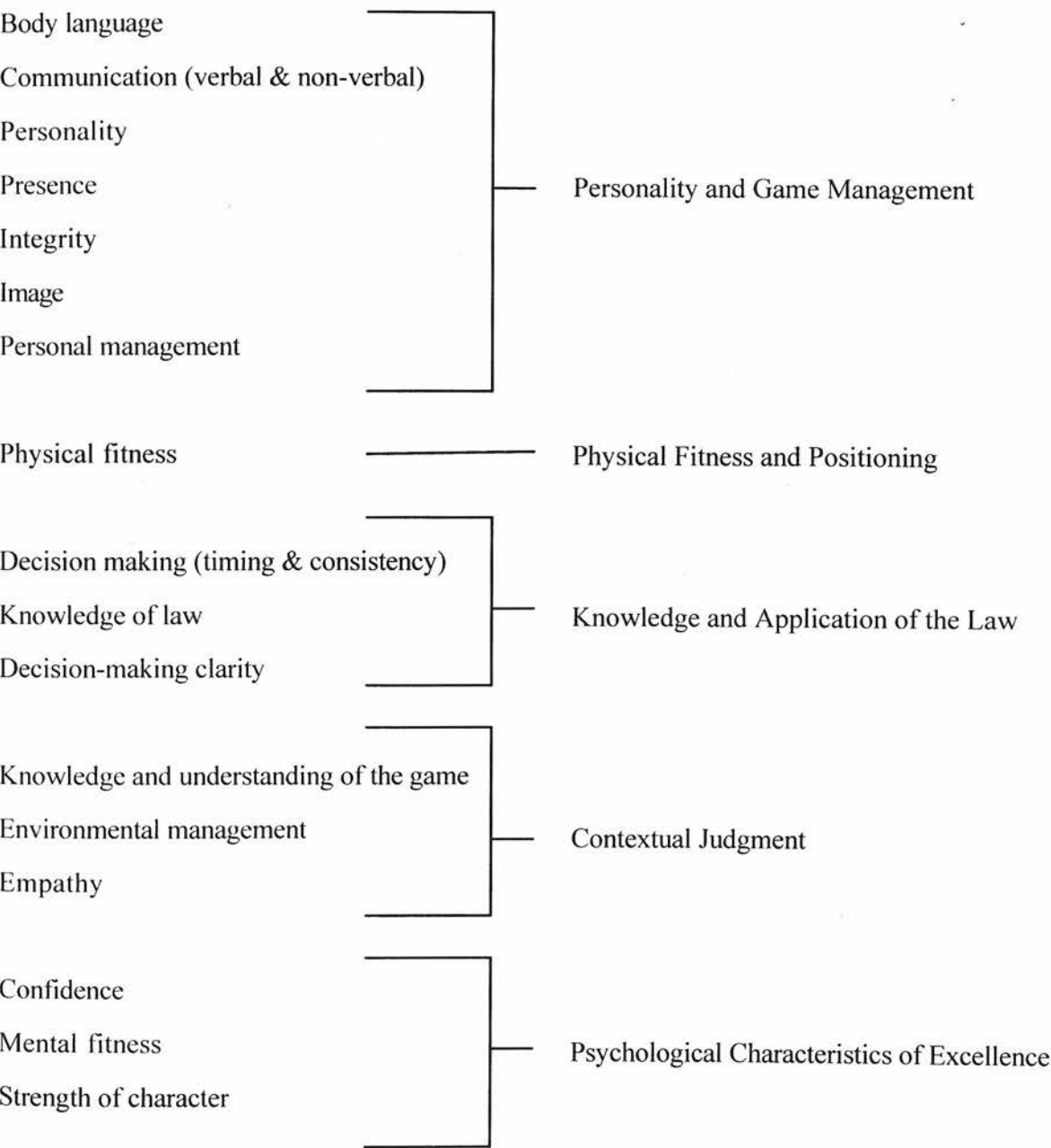


Figure 2.2 - Performance Profiles



## *2.44 Empirical Research*

The review of empirical research showed several recurring themes. Almost half of the articles focused on referee DM and judgments, with 14 of these 25 studies exploring biases such as crowd influence and home advantage (e.g., Nevill et al., 2002), the effect of players' aggressive tendencies (e.g., Jones et al., 2002), normative rulings (e.g., Rainey & Larsen, 1998), and contingency effects (e.g., Plessner & Betsch, 2001). Three papers examined the effects of officials' positioning on their decisions (e.g., Oudejans et al., 2000) and eight looked at the efficacy of decisions and judgments (e.g., Craven, 1998). Other common themes emerged, including investigations into the physiological demands of refereeing (e.g., Krustup & Bangsbo, 2001) and as discussed earlier many articles explored psychological stressors, focusing on sources of anxiety, stress, coping, and assaults on sports officials (e.g., Anshel & Weinberg, 1999).

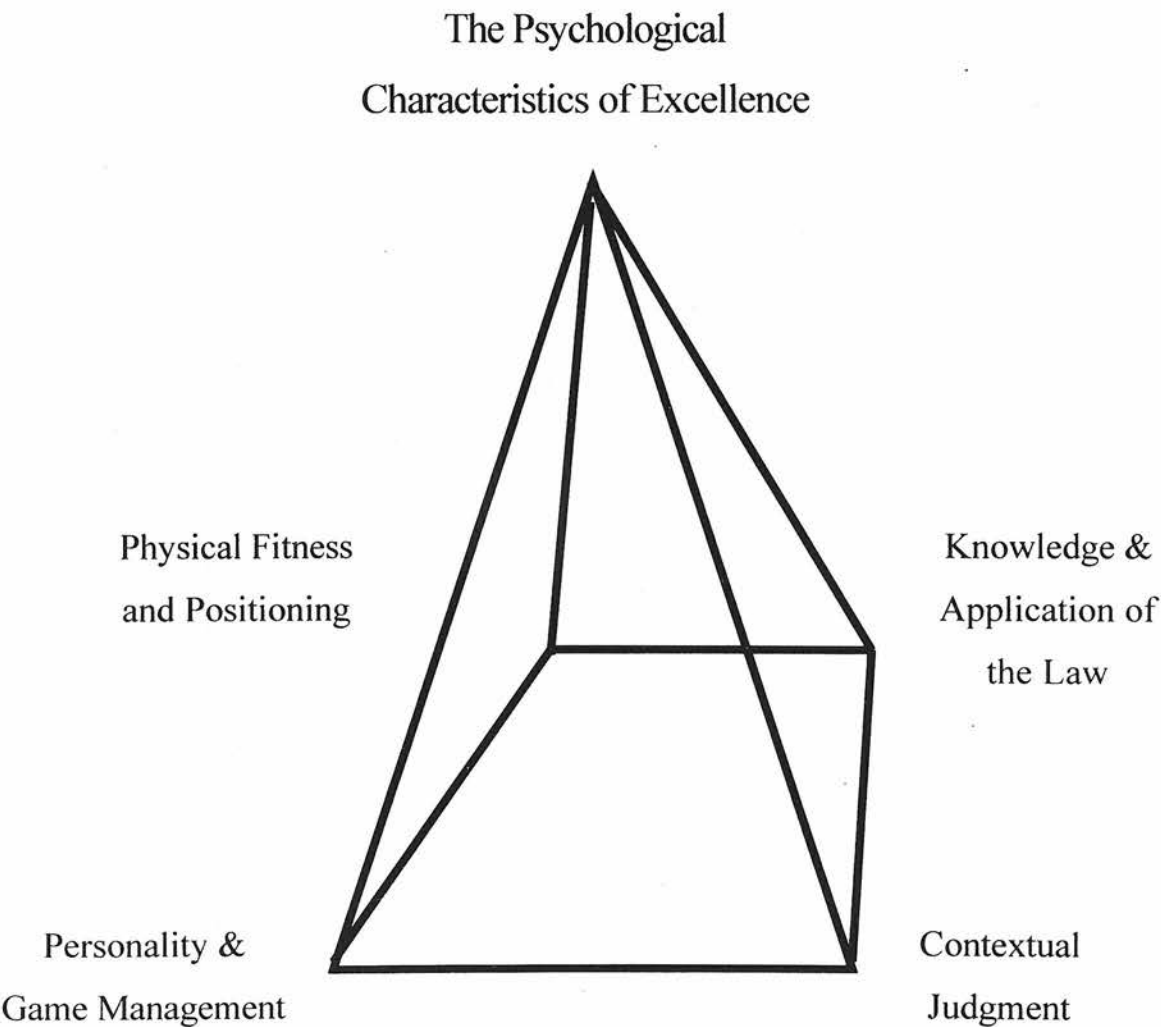
Unfortunately, as previously noted, many of this final category of research themes appeared to have used refereeing as the medium through which to explore psychological phenomena, rather than examining the factors that directly affect refereeing performance. Thus, despite some interesting findings, in cases where a psychological issue rather than performance was the dependent variable, such topics provided little salience to the understanding of performance characteristics. Indeed, stress-related issues were not apparent in any of the data yielded by the other two sources.



This feature, researching through refereeing rather than for it, may provide an explanation for the failure to find empirical research that specifically examined communication skills in sports officials, despite the overwhelming evidence from all the other sources suggesting that it is a crucial element of the job (e.g., Bunting, 1999; Melrose, 1999). Given such circumstances, the “psychological stressors” component was not followed up, especially since skills such as coping were adjudged to be subsumed within the component of “psychological characteristics.” The all-purpose nature of such psychological characteristics deems them crucial for development in all aspects of performance (Orlick, 2000) including refereeing, and as such, it was felt that they should overarch all the other areas. For example, referees who have no commitment or cannot set goals will be limited in their capacity to make a positive change in any of the key areas. Hence, the “Cornerstones Performance Model of Refereeing” emerged, featuring four base areas of performance and overarched by the psychological characteristics of excellence (see Figure 2.3).

Given very specific comments in the RFU literature and subsequent characteristics identified in the performance profiles, it was felt that the characteristics were best reflected as a model to indicate the relationship between the cornerstones. For example, although “empathy” might be based upon good contextual judgment skills, it may also be manifest in personality and management skills and could even impact upon the application of the law.

Figure 2.3 - The Cornerstones Performance Model of Refereeing



## 2.5 Phase 2 Validating the Model

To validate the model it was presented to three independent groups of individuals representative of elite refereeing. These focus groups included a group of high potential referees who had been placed on a fast-track program, then ranked between 50 and 65 on the English national panel ( $n = 6$ ), the three top ranked, full-time referees ( $n = 3$ ) all of whom were also active at international level, and the full-time management staff at the RFU referee department ( $n = 7$ ). The latter group comprised both development officers ( $n = 4$ ) and referee development managers ( $n = 3$ ). I purposely studied developing referees, established referees and the individuals responsible for training both groups to gain a balanced perspective on the efficacy of the model. Individuals were given 5 minutes to independently critically analyse the model and the definition for each cornerstone (see Appendix One), before discussing their comments amongst the group for a further 15 minutes. Specifically, they were asked if it accurately reflected all the key aspects of referee performance and to seek out improvements to the model.

## 2.6 Results and Discussion (Phase 2)

All three groups universally accepted the model as isomorphous to refereeing performance and as a valuable aid to referee development. High potential referees felt that it offered a clear focus for breaking down their performance. The full-time referees also believed it to be representative of refereeing rugby union and highlighted that knowledge and understanding of the game formed the key underpinning to

contextual judgment. The RFU referee management staff was equally supportive of the model and felt that recognising the relationship between various cornerstones was critical in determining such things as the “X-factor of refereeing.” In fact, this group felt that contextual judgment’s interaction with both psychological characteristics, and personality and management skills went a long way towards clarifying this X-factor, which they had previously been unable to define yet deemed critical for high-level refereeing.

The only concerns expressed were in making various sections of the model operational. For example, the RFU management staff felt that physical fitness and positioning together with law application skills would be considerably easier to train than contextual judgment and personality and management skills. After detailed discussions, the management team suggested that these more easily trained areas represent the “robotic” skills of refereeing in that there is a given standard of application, whereas the other more amorphous cornerstones, contextual judgment and personality and game management skills, represent more “humanistic” elements of performance. Nevertheless, despite the suggestions that some cornerstones might be easier to train than others, the model was accepted unanimously as a valuable tool for examining referee performance.

## 2.7 Phase 3 Using the Model and General Discussion

As a result of this investigation the RFU adopted the Cornerstones Performance

Model of Refereeing to guide training for their elite referees. Accordingly, the model was the driver for the R&D based support services offered to the RFU throughout the project and was used as the basis of in-service training, coaching, talent identification and development, as well as for assessing referees to be put forward for iRB duties. Reflecting my earlier comments, the benefits of the model's development exceeded the direct structural considerations, offering a "multiplicative" evolution through increased shared thinking and common goals. Reflecting the precepts of NDM (see Chapter Three), all involved were able to clearly communicate, explore and expand on crucial concepts, facilitated by a common language. Furthermore, all were able to move forwards secured by greater confidence in the veracity and impact of the ideas they were working with. Finally, since its development, the model has received support from other bodies both in the UK and USA, such as the Rugby Football League, US Major League Soccer and the National Association of Sports Officials (e.g., Mascarenhas, Collins & Mortimer, 2002b).

One of the key strengths of this investigation over previous research was that the sample of referees studied was truly elite, including several international and ex-international referees. As highlighted in the first chapter most of the literature on sports officials has investigated samples of referees merely described as qualified. This chapter highlights findings that may be more reliably generalised to elite officials responsible for arbitrating professional sports. Also, in triangulating the methods (content analysis, performance profiling, and interviews) and sources (different

groups of referees, referee development officers and data from assessors), together with “member checking,” the findings were more likely to present valid results (see Krane, Anderson & Streat, 1997). In addition, it is important to recognise that the themes developed here represent a starting point for support work; in short, a dearth of effective empirical data on the performance environment should direct practitioners to this type of approach.

## 2.8 A Systematic Approach to Examining the Model

With a performance model established, the sport science team began to provide support in each of the cornerstones. Concerns over how to make the model operational have been addressed. For example, to quantify the physical demands of refereeing, the sport science team completed a notational analysis of the movement patterns of premier league referees. Replicating the patterns associated with premier league refereeing, a rugby union referee specific fitness test was developed (Martin, Tolfrey, Smith, & Jones, 2002) to formally assess referee specific fitness.

Development of the “psychological characteristics of excellence” and “personality and game management” skills were largely conducted through the referees’ access to individual support, in much the same way as any group of elite athletes would use such a service. This is described in greater detail in Chapter Seven. The physical fitness cornerstone was further developed through access to the team’s physiologist, who provided personalised fitness assessments and tailor made training programmes.

Subsequently, the two remaining and arguably most important cornerstones (Anshel, 1995), knowledge and application of the laws and contextual judgment became the focus. As highlighted earlier, the aim was to *understand* each cornerstone, find a method to *operationalise* it and finally develop interventions to *train* it.

## Chapter 3. Reviewing the Decision-Making Literature

### 3.1 Preamble

Given such emphasis from NGBs and the development of a performance model which suggests that, at the very least, 40% of the task is based on DM (2 of the 5 corners of the model), there appears to be a consensus that DM is the most crucial factor for referee performance. Therefore, evaluating DM skills would seem to be imperative for increasing the quality of officiating. Accordingly, it is necessary to turn to previous DM research in the field of sport science to find an appropriate approach to explore this crucial area.

### 3.2 The Demands of Rugby Union Refereeing

Anshel's (1995) survey of referees suggested that knowledge and application of the laws of the game represents the key skill for a referee (Anshel & Webb 1991) and as such would appear to be the prime candidate for a psychological performance enhancement intervention. A variety of different approaches have been used to enhance the quality of officials' DM during games. For example, video-technology has been used to assist officials in cricket, rugby-league, rugby union (iRB, 2003) and American football. The application of such techniques implies that a referee should be an automatic and robotic dispenser of the laws. For example, the National Hockey League has recently offered exact descriptions of specific penalty infringements and officials are constantly checked to ensure that they are applying the laws in the required fashion. The same is true for association football referees



and their assistants who have very clear interpretations and guidelines on specific laws, such as the offside law for refereeing in the Football Association Premiership.

However, as highlighted in Chapter One, few sport situations will ever be so simple that a clear “if...then” decision is easily applicable and it is likely that some element of human judgment will always be required, if only to maintain the dynamic flow of a game. For example, on occasions the referee is actually encouraged to not call or sanction a foul, or perhaps sanction it differently in different circumstances as part of the essential skill of “managing the game” (FIBA, 2004; Grunski, 1999). In fact, it is this capacity to apply the laws within the spirit rather than the letter of the law (termed contextual judgment, see Chapter Two), which seems to be a primary requirement for elite performance as an official. Hence referees who do not make decisions in accordance with NGB guidelines on the laws, not only thwart their progress as referees but in some cases may face relegation to refereeing lower level games (e.g., The Guardian, 2001). In light of such emphasis on the referee’s ability to read the game, I felt it necessary to first explore literature on visual perception in sport, as correct decisions may only be made if the correct information is first identified.

### 3.3 Perception and Vision Lessons from Motor Control Literature

Refereeing is about perception leading to DM (Ste-Marie, 2003). Although typical motor learning studies focus on performance where perception leads to action, motor

control literature has also investigated many pertinent perception issues that refereeing is challenged with. For example, research into experts' cue-utilisation in sport has lead to some valuable findings on the their differential fixation and search strategies (see Williams, Davids & Williams, 1999). Likewise, with regard to sports refereeing, knowledge of the pertinent areas of the situation is critical in order to allocate attention appropriately.

On reflection of such, motor control studies have used event occlusion methods to identify the critical cues in the task that experts attend to, in order to support an appropriate response. Much of this research has explored single actor scenarios performing relatively simple skills, i.e., subjects attending to only one performer, such as a penalty kicker in soccer (e.g., Williams et al., 1994) or badminton player performing an overhead shot (e.g., Abernethy & Russell, 1987). The environment in which team sport officials operate is rarely that straightforward, however. Very dynamic situations have to be monitored, often with simultaneous events occurring on different parts of the field. This complexity is likely to preclude such formulaic findings to the development of open team sports officials. Interestingly, findings so far suggest that even in a relatively simple skill like a football penalty kick, blanking out different areas of the penalty kicker led to experts using compensatory strategies that proved to be equally successful (Williams, Davids, Burwitz & Williams, 1994). So it appears that a prescriptive strategy might not always be appropriate, particularly when the added complexity of rugby union refereeing is the area of examination. In fact, expert decision-makers such as rugby union referees are able to make appropriate decisions even without significant information, yet they are equally

accurate in their DM even when unsighted or presented with ambiguous information; a process that has been termed “satisficing” (Simon, 1955; Stokes, Kemper & Kite, 1997). Thus, developing knowledge and understanding may be more important than training cue-utilisation skills.

Furthermore, whilst sports performers, particularly in racket sports, are taught deception skills in order to mislead their opponents into misreading a shot, it is reasonable to assume that professional team sports players will adopt similar strategies in order to mislead referees (see *The Observer*, 2002). Therefore, incomplete and misleading information may well be an occupational hazard of the environment that the referee regularly encounters. Also, as conceded by Abernethy and Russell (1987) visual fixation may be helpful, but the importance of fixation in a task that has widespread cues may only reflect a pivot point for peripherally orientated visual perception, rather than centrally orientated perception.

Thus, while cue-utilisation theories may be helpful in some environments, in the complex arena of refereeing open team sports, it may be that sport specific knowledge is the critical factor, superseding the need for complete information. This seems to concur with Ste-Marie (1999), who observed that highly skilled players are better able to understand advance information to predict the outcome of visually presented information. So for example, a referee who is unable to see where the ball is in a tackle situation, where such information is crucial (see law 15 iRB, 2003), is still able to make an accurate decision based on a process of “filling in the gaps,”

through the experience of encountering similar situations previously. The process of making such decisions leads to research on judgments made by sports officials.

### 3.4 Lessons from Judgment Literature in Sports Officiating

Ste-Marie (2003) offers a very useful synopsis of judgement literature and its relevance to refereeing team sports, contending that there are both similarities and differences. Firstly, and analogous to the perception-action literature, judgments are often based on single actor events involving comparisons to a perfect model from an external perspective, whereas open team sports officials are a dynamic part of events that may contain up to 30 players at a time. Pressures are high for both judges and referees but the speed of decisions and visibility of performance is generally much higher for the team sports referee. In addition, judges' performances are discrete rather than continuous and interactive in nature, and as such there is a flowing interaction between performers and the referee from one decision to the next. Accordingly, an analysis of DM theories would seem to offer a more holistic approach.

### 3.5 In Search of Appropriate DM Theory

#### *3.51 Classical Decision Making.*

According to classical DM theory, a referee's application of the laws would involve identifying the problem, generating a range of alternative solutions, evaluating these

solutions, and finally selecting the preferred option (Cannon-Bowers, Salas & Pruitt, 1996; Beach & Lipshitz, 1993). However, although classical theories have been successfully applied to structured problems where the speed of decision is less critical, applying the theory from these situations to highly time pressured, demanding, ambiguous, and dynamic situations such as refereeing is inappropriate and rarely occurs in the real world (Lipshitz, 1993).

Referees typically take under 1 second to make a decision (Jones et al., 2002) and given the physical and cognitive demands of refereeing, together with such temporal pressures, developing a range of alternative solutions, systematically evaluating them against one another, and choosing the most appropriate course of action would intuitively seem to be an inappropriate strategy. Rather, Klein (1993) would suggest that referees use recognitional strategies, making decisions based on prior knowledge and experience. In fact, Klein's (1989) "Recognition-Primed Decision" model that propounds this method of decision-making was based on empirical research with fireground commanders, who although under considerable pressure, have far less *time pressure* than referees. Furthermore, none of the published empirical research papers to date, that has explored referee decision-making has either propounded or adopted a classical DM perspective.

Importantly, when classical decision theories have been taught to experts such as managers, they rarely use them (Beach & Lipshitz, 1993), choosing to ignore the prescriptions that oppose their own subjective intuitions (Dreyfus & Dreyfus 1980; Isenberg, 1985; Roth, 1997). Beach and Lipshitz (1993) suggest that these

(successful) executives find the prescribed operations cumbersome, time consuming and inefficient on their time and resources.

Given the experts' distaste for rational choice strategies, one might assume that the time constraints force them into some form of speed-accuracy trade-off. As demonstrated in a series of gambling scenarios (Payne, Bettman & Johnson, 1988), subjects were found to completely change their DM strategies when asked to perform under increasing time pressure, yet surprisingly studies of chess grandmasters and experienced aircraft pilots found that the expert's accuracy loss was minimal when decision speed was increased (cf. satisficing). For example, chess grandmasters, when confronted with a complex situation often discover the correct move to make within 5 seconds or less but then might take up to 15 minutes to "internally verify" the correctness of this move (de Groot, 1965). More recently, Calderwood, Klein and Crandall (1988) found that the proportion of blunders made by chess masters, as rated by an independent grandmaster, showed no effect of time pressure with error rates of 8% under regulation time (approximately 2.6 minutes per move) but only 7% under blitz conditions (moving within 6 seconds). In contrast, the proportion of errors made by class B players did increase with time pressure from 11% to 25%. Similarly, Stokes et al. (1997) reflecting on experienced pilots' ability to chose the correct answer as their first choice 71% of time, compared to the less experienced pilots whose first choice was only correct 53% of the time, concluded that the experts' ability to make the correct decision first time without generating and deliberating alternatives was achieved through the use of knowledge representations

or situational schemata from long-term memory, rather than the techniques propounded by traditional approaches.

### *3.52 Naturalistic Decision Making*

So, given such evidence, rather than prescribing the process by which DM should be conducted, researchers have turned to studying the methods that the experts actually use in the real-world by exploring expert military commanders, aviators, fire-fighters and astronauts on the NASA space program. Researchers found that when confronted with these high-stress environments and when decisions must be made quickly, experts are able to make effective decisions without the need for deliberation over a range of alternatives (Klein, 1997a; Orasanu, 1993, Zsombok, 1997). This new research approach, naturalistic decision making (NDM; Orasanu & Connolly, 1993) offers an extremely effective perspective for the examination and enhancement of DM skills that is more suited to the naturalistic environment in which sports officials operate.

The growing body of research into NDM has led to a clearer definition of this perspective; characterised by ill-structured problems; uncertain dynamic environments; shifting, ill-defined or competing goals; multiple event-feedback loops; time constraints; high stakes; multiple players and organisational norms and goals that must be balanced against the decision-maker's personal choice (Cannon-Bowers, Salas & Pruitt, 1996; Zsombok, 1997). Although it has become generally accepted that only some of these factors need to be present for a decision to be considered naturalistic (Orasanu & Connolly, 1993), all nine are characteristics of the

refereeing environment. Refereeing often presents complex and unstructured scenes with players moving at high speeds, where a referee has flow and control issues to balance, making decisions that are not isolated but wrapped into evolving patterns of play. They are expected to react instantaneously often under intense media scrutiny, while accounting for the input from other refereeing team members (in rugby union's case the touch-judges, who control the sidelines and have a microphone link to the referee). They also have to balance their interpretation of law with the organisation's philosophy on the way they feel the game should be officiated. Consequently, the NDM research paradigm, pioneered in very similar "high stress" environments with extreme time pressure and where poor DM may have serious consequences seems to offer a suitable framework in which to study refereeing performance in a team sports environment (McLennan & Omodei, 1996).

In this team sport environment, as identified by Brehmer (1992), three types of DM styles exist; (1) command based, where each decision is made centrally and relayed to the other teammates; (2) by plan, often referred to in the military as "Standard Operating Procedure," where teammates act to a prescribed format with no need for communication; or (3) by experience, where teammates respond according to the nuances of the situation and the decisions made by teammates.

An example of command based DM in sport is the quarterback in American football, who calls the play to his teammates who then carry out their roles accordingly. A free-throw in basketball exemplifies a plan based DM style in sport, where each defender lining up along the free-throw lane "boxes-out" the player next to him or



her (i.e., puts their body in front of the attacker to prevent them from rebounding the ball), and the defender at the top of the key boxes-out the shooter. In this example, there is no need for communication as this is simply the standard operating procedure when a free-throw is taken. Finally, experience based DM is built upon very detailed knowledge of the task, the team and the environment (Kraiger & Wenzel, 1997), which allows team members to function synchronously. This can regularly be seen in open team sports such as soccer, when players are able to anticipate the movements and actions of teammates and adjust accordingly, resulting in fluid and implicit interactions that lead to faultless passages of play (Riley, 1993).

This NDM approach also uses a variety of different and more ecologically appropriate techniques to examine DM within the context in which it occurs; a factor which sport science literature has long been advocating (e.g., Williams et al., 1999). Indeed, preliminary NDM findings suggest that in high time-pressured settings, expert decision-makers use *pattern matching* or *framing*, based on mental models rather than the systematic judgment of the losses and gains of self-generated solutions; a discovery that would seem to concur with motor control research in sport where more skilled performers were found to rely upon their superior recognition strategies for typical patterns of play (e.g., Williams et al., 1994).

NDM investigations have been applauded for their attention to the study of experts, leading to considerable success “in the field,” with interventions in domains such as military command (e.g., Noble, Boehm-Davis & Grosz, 1986), aviation (e.g., Schneider, 1985), fire fighting (e.g., Klein, Calderwood & Clinton-Cirocco, 1986),

emergency nuclear power plant management (e.g., Roth, 1997), and naval command (e.g., Hutchins, 1997). Furthermore, as the area is still developing, it provides an excellent opportunity to advance new applied research methods and adapt new theory to a pertinent population of experts.

### 3.6 The Theoretical Underpinnings of NDM –

#### How Do the Experts Do It?

Since NDM has only received recognition in the last 15 years, most of the work has developed from intuitive insights into observations of experts, leading to a variety of theories. Interestingly, many of these theories appear to depict the same processes which have merely been described under different labels. Importantly, while many of these offer plausible explanations of experts' decisions, one must adopt a single approach in order to ensure clarity and coherence. Underpinning all NDM theories is the assumption that the expert exhibits superior knowledge, which is organised in such a fashion that it can be accessed instantly in order to guide decisions made under extreme pressure (Druckman & Bjork, 1991).

Attempts to measure those knowledge structures that underpin performance have highlighted some important concepts. Anderson (1995) suggests that knowledge can be categorised into either declarative or procedural. Procedural knowledge tells us “how to” perform a task, whilst declarative knowledge represents the “whats and whys” of that task. For referees, the declarative knowledge represents the knowledge and understanding of the players, laws, environment, and the nuances on the day, and

the procedural knowledge is seen in the tools the referee uses, such as his or her use of the whistle, voice, and body language.

Given this pressure, researchers have found that skilful decision-makers develop an ability to *chunk* (Means, Salas, Crandall & Jacobs, 1993) a decision into a *type* (Cannon-Bowers & Bell, 1997) with highly familiar patterns. Tversky and Kahneman (1981) call this process *framing*, suggesting that it helps trainees to see situations in the same way as experts who parse the pattern of cues rapidly through *pattern recognition* (Means, et al., 1993) or *pattern matching* (Klein, 1997a) into a *template* (Cannon-Bowers & Bell, 1997). Despite the potential source distortion in the categorisation of decisions into themes (Tolcott, Marvin & Bresnick, 1989), researchers sanction it as an effective way of organising information when time pressure is inherent.

However, despite overwhelming support for some form of knowledge organisation, as this involves the classification of meta-cognitions there is very little empirical research that rigorously explores the precise manner in which this knowledge is classified. Nevertheless, many of these theories build upon very well established constructs throughout psychology literature. The most recognised of all these emerging theories, which has also received considerable support outside the NDM literature, suggests that experts use *mental models* to help team coordination and performance by forming accurate and meaningful expectations and explanations of events (Cannon-Bowers, Salas & Converse, 1990; Norman, 1983). In this regard, Rouse, Cannon-Bowers and Salas (1992) suggest that mental models provide

knowledge of the situation to direct attention, classify information and to form an understanding of its meaning. More recently, Endsley (1997) highlights that mental models create expectations regarding the future status of events and provide a link between recognised situation classifications and typical actions. Thus, they are not only capable of directing attention but also triggering potential solutions.

The purpose of mental models is to aid interpretive processes, helping to screen out information in order to prevent overload; a factor that seems crucial when individuals are presented with extreme time pressure. Holyoak, (1984) defines mental models as “a psychological representation of the environment and its expected behaviour” (p. 193), thus helping individuals to make sense of their surroundings (Klimoski & Mohammed, 1994). They are used to both represent and organize knowledge by simulating the important characteristics of situations that are to be cognitively mastered, and they do so in such a fashion that complex phenomena become plausible (Seel, Al-Diban & Blumschein, 2000).

Therefore, learning a complex cognitive skill can be achieved through developing mental models, as they describe both the procedural and declarative knowledge that is required for effectively solving problems at each stage of acquiring the skill (Seel, Al-Diban & Blumschein, 2000). They are formed on the basis of construction, revision, and evaluation of real-world situations (Johnson-Laird, 1983). Thus, exposing individuals to a variety of real-world scenarios will help them to create mental models of those events, as evidenced by Seel et al. (2000). Seel and his colleagues found that the generation of mental models on the basis of former

experiences with similar situations, helped individuals to master new situations. Accordingly, they are a helpful training aid for referees who regularly experience new situations that are similar to those that they have experienced before (Mascarenhas, Collins & Mortimer, 2005).

Mental models also embrace a variety of analogous concepts such as *schemata theory* (Rouse & Morris, 1986) and *knowledge structures* (Druckman & Bjork, 1997). In addition, since mental models provide a conceptual framework for describing, explaining and predicting future system states (Rouse & Morris, 1986), this construct seems to incorporate aspects of situation awareness.

Situation awareness advocates three levels of awareness; (1) perception, (2) comprehension, and (3) prediction, which lead the decision-maker to an appropriate decision (Endsley, 1997). In fact, it is this third level of awareness that seems to allow team members to predict the actions of others and adjust their performance accordingly. The shared nature of such knowledge is worthy of further exploration, since both interdependent and independent understanding is critical to referee DM performance. Therefore, although mental models have traditionally been examined at the individual level, recently there have been considerable efforts to explore these phenomena at the group level (Klimoski & Mohammed, 1994).

### 3.7 The Development of Shared Mental Model Theory

The naturalistic study of DM postulates that when team members have accurate and equally detailed conceptualisations of the problem and the requirements of team

functioning necessary to solve it, they have developed shared mental models (SMMs - Cannon-Bowers et al., 1990; Noble, Grosz & Boehm-Davis, 1987). The earliest reference to the maintenance of a *shared* mental-model was in Torrance's (1953) report on World War II military units downed behind enemy lines. In this account, the leader simply took responsibility for communicating information to develop a shared understanding of the situation; more recently, and perhaps more appropriately, this has been termed a *shared situation-model* (see Rouse & Morris, 1986). Although the constructs in Torrance's research are not the same as the SMM construct it does seem to match theories on problem *framing* that appear to drive the construction of appropriate mental models. Since then, Heath (1991) described how observations of a baseball team over a season exposed and developed SMMs, thus reducing the need for lengthy and explicit communication between team members, which they had previously required.

SMMs provide team members with the ability to predict both future events and the needs of other team members. This allows inferences to be drawn, in order to anticipate teammates' actions and make adjustments to maintain team coordination (Johnson-Laird, 1983; Rouse, et al., 1992). SMMs guide the evolution of "same expectations" and "same patterns of thinking" that lead teammates to all respond in the same manner (Cannon-Bowers et al., 1990). Furthermore, they are postulated to provide the knowledge, attitudes and behaviours to form a common understanding of a situation. When two or more individuals share this level of understanding, based on the same recognition of the situation, comprehension of the problem, and prediction of future events, it often leads to the same decision outcome and they can be said to

be *coherent*. Thus, knowledge of the various components and the relationships among them, provided by a SMM, leads to coherence (Rouse & Morris, 1986; Rumelhart & Ortony, 1977).

Crucial for the purposes of this research, this collective understanding promotes a similar appreciation of the situation, and seems to provide the platform for decision-makers to *independently* judge situations coherently. Thus, if coherence is a measure of a team's DM prowess, assuming that decision-makers are conforming to *accurate* decisions, then the promotion of more effective DM is contingent upon the development of their collective mental models of the situation (Lipshitz & Ben Shaul, 1997). Hence, the extent to which these referees share appropriate mental models will dictate the extent to which they interpret situations concordantly, a crucial feature of refereeing performance.

### 3.8 Naturalistic Approaches to Training Coherent DM

Currently, referees appear to rely on gaining refereeing experience to develop expertise. Unfortunately, simply amassing experience does not necessarily lead to expertise (Williams & Davids, 1995). Refereeing alone may not provide a sufficient number of varied and challenging scenarios in close succession to develop expertise through the mere experience of refereeing (cf. Means et al., 1993). Thus, since effective teams need a number of accurate mental models to form an understanding and generate predictions about events (Salas, Cannon-Bowers & Johnston, 1997), presenting "typical scenarios" to trainees may be an attractive alternative to on-the-job learning (Stokes, et al., 1997; Ericsson & Lehman, 1996). Means et al. suggest

exposing trainees to a *variety* of scenarios to improve pattern recognition and form multiple mental models of typical, “representative” situations, using cues and patterns that characterise the specific performance domain, as a solution to the training challenge (see Cannon-Bowers & Bell, 1997; Cohen, Freeman & Thompson, 1997).

Through such naturalistic training, trainees should not only learn the cognitive skills required to make rapid and accurate assessments of situations through an understanding of the significance of certain cues, but they may also acquire the meta-cognitive skills by learning how to allocate their mental resources effectively. For example, Zakay and Wooler (1984) found that training without time pressure did not enhance DM under time constraints and when student aircraft pilots have been exposed to such pressure training, subsequent in-flight stress was reduced (Krahenbuhl, Marett & Reid, 1978). In the sporting domain, there is evidence to suggest that novice squash players can improve their anticipatory skills through structured and knowledge-based video training (Abernethy, Wood & Parks, 1999). Furthermore, the ability to use such perceptual and cognitive skills has been shown to delineate between elite and sub-elite association football players as early as age nine (Ward & Williams, 2003). Thus, the potential positive training effects for sports officiating, a task that is arguably far more cognitively biased than sports playing, are significant.



### 3.9 Video Scenarios to Train DM Performance

Video scenarios provide trainers with the flexibility to present clear and timely feedback, rather than the more typical analysis, which may be sometime after the incident under scrutiny has occurred (Means et al., 1993). Additionally, practice problems can be designed specifically to challenge the individual decision-maker at their current level of expertise (Means et al., 1993).

Video has been used to assess DM perspective (McLennan & Omodei, 1996), with structured video training providing a useful adjunct to “in-game” training (Starkes & Lindley, 1994). Although Starkes and Lindley (1994) failed to show any conclusive evidence of such training transferring back to the game, other researchers suggest that transfer will occur when transfer-appropriate processing occurs. That is, transfer will occur to the extent that the scenario demands the same cognitive processes that are required in the real task (Starkes & Allard, 1993). Furthermore, investigations into social learning theory (Bandura, 1977) and subsequently observational learning research (McCullagh & Starek, 1999) suggest that attention, retention and rehearsal are also crucial to ensure that the new skill is developed and transfers to the real-world environment. Therefore, having found the most appropriate theoretical approach, it was necessary to find an area of the game suitable for investigation.

## 3.10 Developing DM in Rugby Union Refereeing:

### The Starting Point

In the specific context of rugby-union the diverse and dynamic environment of the tackle (law 15) would provide a uniquely demanding task to accurately interpret the coherence of DM (E. Morrison, personal communication, July 5, 1998). The tackle regularly creates the most controversy and is thought by many to be one of the most complex events to referee in all team sports (Ackford, 2003; Bunting, 1999). It presents a unique situation where multiple, complex and dynamic decisions are required as there are timing elements, overlapping elements, interactive elements, and often multiple players involved in the action (see Ackford, 2003).

Support for investigating this law was shown in a study of English premier league and national league (England's top divisions) coaches, in which 87% of respondents felt that communication with referees throughout the year should include interpretation of this particular law (Melrose, 1998). Therefore, building on these various practical considerations and theoretical contentions, the aim of this phase of the programme was to construct tools capable of measuring and training the key aspects of referee performance. In addition, if successful, I felt that it would provide a preliminary investigation into the use of SMMs in refereeing.

# Chapter 4. He May Not Always Be Right, But He's Always the Referee: A Method to Assess the Accuracy and Coherence of Rugby Union Refereeing

## 4.1 Preamble

Having established an appropriate theoretical approach, the crucial first step was to empirically explore a method to assess the most important cornerstone, knowledge and application of the law (Anshel, 1995). It was anticipated that this would give the RFU a method to highlight areas of concern and identify any performance disparities between different groups of RFU officials. Furthermore, it was expected that findings would inform the development of future training interventions in this area.

## 4. 2 Introduction

As discussed previously, a naturalistic approach was adopted since rugby union demands rapid DM, requiring referees to evaluate the important characteristics of an event and present an appropriate solution in about 1 second (Jones et al., 2002), without the opportunity for reassessment or contemplation on the implications of their decision. Referees have to respond quickly to dynamically unfolding events, which may hold many uncertainties and ambiguities, and often in response to input from touch-judges (the officials responsible for controlling the sidelines in rugby union who have a microphone link to the referee).

Given such challenges an ecologically valid but scientifically rigorous approach was required, matching the precepts of NDM. Usefully, previous NDM investigations have studied the efficacy of DM assessment and training methods (e.g., Stout, Cannon-Bowers, & Salas, 1996) concluding that they need to be of sufficient functional quality to test the experienced decision maker's ability (Alessi, 1988; Klein, 1997a). In this context, video and audio presentations provide a suitable format for assessing such DM skills (Abernethy, 1996; Cannon-Bowers & Bell, 1997; Williams & Grant, 1999). Similarly, Omodei, McLennan and Whitford (1998) suggest that "own-point-of-view" video recordings can provide the best representation of the complexity and dynamics of naturalistic environments and in addition, allow the selection of pertinent events from a wide variety of data. Unfortunately, despite the evidence supporting video as a medium through which to assess and train DM, there is no empirical research that has examined suitable criteria to measure relative success in referee DM performance.

#### *4.21 Measuring DM Performance – Accuracy, Coherence and Shared Mental Models*

Rugby union refereeing requires referees to make decisions that are not only accurate in law but also consistent with NGB interpretations and in harmony with other refereeing peers. Hence, the correctness of a decision (at least for players and coaches) is substantially based on recent regular experience of the performance environment. Although the published laws are clarified and interpretations may be guided by advice from the governing body, it is the coherent application of the laws

by the senior referees that provide the benchmark for success (Bunting, 1999). As noted previously, examinations of umpires calling “force-out” plays in baseball would seem to support this, as they were found to collectively adopt a *normative* rule in their adjudication of “phantom tags” (Rainey & Larsen, 1988; Rainey, Larsen, Stephenson, & Olson, 1993).

Thus, performance evaluation needs to consider the individual’s DM against his or her peers rather than an independent body of experts, a factor which accounts for the different styles of refereeing across the rugby union world, despite the fact that the iRB laws govern all games. The Tri-Nations’ league known as the “Super 12” has evolved into a far more liberal interpretation of the laws when compared to the English premier league (J. Barnard, Top 10 national ranked referee, personal communication, July 3, 1999; Jones, 2005). Therefore, group consensus is an important factor to identify since it serves as the criteria of correctness, providing measures are taken for both the outcome and the reasoning that has guided the DM process (Mascarenhas, Collins & Mortimer, 2002a; see Abraham & Collins, 1998). Only then, when referees share the same expectations, assumptions and DM strategies can they be considered coherent in their application of the law (see Millgram & Thagard, 1996). This is critical, since mere conformity does not necessarily mean coherence. For instance, a team member who makes the same decision as another but arrived at that decision through a different DM process does not necessarily share the same understanding of the situation. Thus, when applying

that incomplete understanding to a different situation in the future, that individual is more likely to make an incongruous decision. For example, in refereeing the rugby union tackle, the tackler is required to “roll away” immediately and then the support players must arrive on their feet. If two referees awarded a penalty to the attacking team for these different reasons, the outcome would be the same. However, the referee who did not notice the tackler’s failure to roll away (which should happen first) has not seen the same thing and may miss this in future tackles. Thus, examining the level of agreement by measuring the range of responses between referee’s *reasons* underpinning their decisions represents an important and face valid criterion for measuring referee performance. Moreover, when decisions are built on a coherent appreciation of an event, teams have the ability to perform together more successfully (Rouse, et al., 1992).

Cannon-Bowers et al. (1990) attributed such coherent performances to shared mental models (SMMs), a concept that serves to explain faultless performance through implicit interactions between members of successful teams (see Brehmer, 1972). Therefore, as a corollary, for rugby union officials these SMMs consist of not only knowledge of the other team members and their roles, allowing effective coordination strategies between referee and touch-judges (who control the sidelines) but also a declarative knowledge base of the task, its concepts, and the relationship between them (Stout et al., 1996). Furthermore, as these SMMs underpin coherent performance by providing similarly organised expectations surrounding the task

(Rouse & Morris, 1986), the development of SMMs can be used as a basis for understanding and enhancing both dependent and independent team DM in real-life settings (see MacMahon & Ste-Marie, 1999; Stout et al., 1996). As such the SMMs of all those involved in the officiating process, the referees, touch-judges, their coaches and assessors, are of interest when exploring the efficacy of rugby union refereeing.

As noted in Chapter Three, given the complexities of the tackle law (law 15) it is likely that a more extensive declarative knowledge base and hence a more complex SMM is necessary for refereeing this area (see MacMahon & Ste-Marie, 1999). Consequently, giving NDM such a rigorous challenge should test the robustness of these methods and also provide implications that will assist DM in other open team sports.

Therefore, the primary aim of this phase was to measure the DM accuracy, agreement and coherence of England's best RFU referees, their assessors, coaches and touch-judges. Specifically, I was interested in the relationship between the officials' accuracy (as measured by their ability to reach an agreed standard), their conformity to each other and the coherence of their reasons underlying their decisions. Secondly, recognising the roles played by different officials in their coherent application of law I was interested in differences between groups. Finally, it was anticipated that the results would highlight specific applied areas of concern in refereeing the tackle and

provide a preliminary application of NDM theories with a video-based system to assess the time-pressured DM of expert officials viewing game scenarios.

## 4.3 Method

### 4.31 Participants

The participants were 132 male RFU officials who were the delegates at the RFU referees national conference. They included 45 of the top 65 RFU referees, 27 referee assessors, 13 referee coaches, and 47 of the top 120 touch-judges. This sample represented 132 of the 239 individuals responsible for either officiating or developing officials in England's top five rugby union divisions. The referees, ranging in age from 27 to 51 years ( $M = 38.6$  yr.;  $SD = 5.6$  yr.) had refereed on the English National Panel from 1 to 16 years. Based on their national rankings (1-65) made by a group of referee development officers in May 1998 from the periodical evaluations of 37 advisors, the referees were already sub-divided into 1 of 3 groups; a top-20 group, who were responsible for refereeing in the premier league (level 1;  $n = 14$ ); a mid-panel group ranked from 21-40, responsible for national league level 2 and 3 games ( $n = 8$ ); and a lower-panel group ranked from 41-65 who officiated at levels 4 and 5 ( $n = 23$ ).

### 4.32 Instruments

To prepare a test instrument, incidents were selected from premier league games, recorded with professional video equipment (Betacam-SP). Each scenario was filmed



in close-up from a raised gantry positioned at the halfway line. Only incidents occurring in the middle of the pitch ( $<20^\circ$  of arc) were examined for inclusion in the assessment tape. This provided a view looking down over the incident, similar to the angle that the match day referee might experience (cf., McLennan & Omodei, 1996).

Further steps were taken to ensure the ecological validity of the test items. From an original tape of 130 tackle incidents compiled from 60 hours of premier league play, an independent expert panel consisting of elite referees ( $n = 4$ ), coaches ( $n = 2$ ) and players ( $n = 2$ ) examined the clips. This group independently graded each tackle on the difficulty of the decision on a three point scale where 1 = easy, 2 = medium, and 3 = hard. In addition, they discarded all the tackles that did not display sufficient information to make an accurate decision or those where they felt the match-day referee's decision would be discernible on the screen. The experts then convened as a group and selected 10 difficult (i.e., grade 3) tackles from those remaining that they regarded as presenting realistic game scenarios for the accurate application of law 15. It was anticipated that the use of difficult yet realistic scenarios would provide information to inform referee DM training in the future. Finally, these 10 incidents were edited together to provide a test instrument.

Each edited clip began with a voice-over that introduced the two teams competing and indicated the team in possession. The tackle incident was then played with approximately 5 seconds of "lead-in" to orientate the participants to the scene. After

the tackle incident the recording cut to black and the title “make your decision now” appeared on the screen.

A response sheet was developed to enable participants to quickly and easily indicate their decisions. This was essential since time pressure, as opposed to slower, more reflective DM, is a crucial factor for naturalistic environments (Klein, 1997b). Participants were given a copy of the response sheet, consisting of a series of boxes in which to indicate their decision, a space to explain the reasoning behind their decision, a Likert scale to rate their confidence in the accuracy of each decision, ranging from 1 (*low*) to 5 (*high*), and a section to comment on the quality of each clip (see Appendix Two).

#### *4.33 Pilot Testing*

Prior to the participants’ assessments pilot testing was conducted using a group of individuals familiar with the rugby laws to verify the qualities of the videotape, suitable viewing positions, the efficacy of the response sheet, and the typical length of time it would take to complete it. Based on this pilot work the following procedure was developed.

#### *4.34 Procedure*

For the purposes of viewing the 10 assessment clips the participants were randomly divided into four viewing groups of no more than 35 (for data-collection purposes only), each having approximately the same number of referees, touch-judges,

assessors, and coaches. The pilot study and subsequent analysis of the results confirmed this to be large enough to minimise variability due to procedural differences but small enough to allow each individual an acceptable view of the screen. They were then informed that their own personal responses would remain confidential and that their results would only be presented as grouped data, depending upon their officiating classification. After the participants familiarised themselves with the response sheet they sat in the darkened room where they could comfortably see the tackle incidents projected onto a screen via a standard VHS video recorder and data-projector. This presented an image about 8 feet wide and 5 feet high. The first clip from the videotape was then played and paused immediately after its completion. Participants were asked to make an immediate decision by ticking the appropriate box. They were then given 3 minutes to complete the remainder of the response sheet and were explicitly told not to change the decision once made. An inspection of the response sheets and observation of participants suggested that all conformed to these instructions.

After responding to all 10 clips in the same manner, participants were asked to compare the quality of information upon which they made decisions here, to the quality of information they “tended” to get as referees on the pitch and write their explanation on the back of the response sheet. This procedure was followed consistently for all four data-collection groups.

#### 4.35 Data Analysis

Two of the full-time RFU referees, at the time nationally ranked 1 and 2, determined the correct response. Replicating the conditions under which the participants were asked to respond, they both independently made their immediate decision on the 10 clips. In cases where these two referees had initially disagreed upon responses (clips 4 and 9) they reviewed the videotape and discussed the clip before agreeing on the most appropriate decision. In fact, their initial disagreement was only minor as in both clips they agreed on which team to advantage but provided inconsistent reasons for such sanctions. For example in clip 4 one expert chose to play on, advantaging the attacking team who retained possession and the other chose to award a penalty to the attacking team. Similarly in clip 9, one expert awarded a scrum to the defending team while the other awarded a penalty. Finally, these experts indicated “how many times per game” they typically had to adjudicate a tackle situation like the one presented in each clip. The expert’s mean frequency ratings (number of occurrences per game) for all 10 tackles was  $M = 10.9$  ( $SD = 7.8$ ).

Participants DM performance was assessed by three measures; (1) accuracy - the percentage of participants achieving the correct decision, (2) agreement - the degree of spread of their responses, and (3) coherence – the similarity of their reasons underpinning decisions. The kappa statistic of agreement ( $K$ ) was used to measure the spread of responses. This offered a ratio of the proportion of times that the raters agreed, against the maximum number of times that agreement was possible, correcting for chance (see Altman, 1991). Thus, a score of  $K = .90$  would represent a “very

good” (high) level of agreement and  $K = .10$  would represent a “very poor” (low) level of agreement as classified by the system proposed by Landis and Koch (1977). In addition to these measures, for each clip the participant’s reasons for their decisions were examined to determine the extent of coherence in their mental models of each event. Similarly, all three analyses were conducted on a group basis, consisting of the three subgroups of referees and the three other support groups, assessors, touch-judges, and referee coaches. Bonferroni adjustments were applied to control for the experiment-wise chances of a type-one error.

## 4.4 Results

### *4.41 Accuracy, Agreement, Coherence and Confidence Levels for All Participants*

Table 4.1 provides details of the percentage incidence of responses made, highlighting the accuracy scores and the kappa statistic of agreement for each clip. The mean level of accuracy across the 10 clips for all participants was 49.6% ( $SD = 28.6\%$ ). High levels of accuracy were achieved for clip 1 (82%), clip 7 (89%) and clip 10 (70%). Naturally these clips also exposed high levels of agreement (clip 1,  $K = .60$ ; clip 7,  $K = .74$ ; clip 10,  $K = .41$ ). In addition, these clips showed very high coherence in the participants’ reasoning for each decision. In clip 1, 95% of the participants who responded accurately showed agreement by awarding the penalty for offside with only 5% penalising for support players arriving off their feet. In clip 7, 94% of the accurate participants awarded a penalty to the attacking team for the defender failing to roll away and similarly in clip 10, 95% of the respondents who made an accurate

decision penalised the ball carrier for not releasing the ball.

#### *4.42 Accuracy, Agreement, Coherence and Confidence Levels by Group*

The mean accuracy scores shown in Table 4.2 revealed that the top-20 referees were the most accurate ( $M = 54.3\%$ ,  $SD = 32.9\%$ ), although interestingly the lower-panel group ( $M = 52.4\%$ ,  $SD = 26.3\%$ ) showed greater accuracy than the mid-panel group ( $M = 47.1\%$ ,  $SD = 28.4\%$ ). Furthermore, despite poorer performance, this middle group of referees showed greater confidence levels in their decisions than all other groups ( $M = 4.4$ ,  $SD = 0.7$ ). The referee coaches were the least accurate ( $M = 43.0\%$ ,  $SD = 37.3\%$ ). In fact, their decisions were less accurate than the referees in 8 of the 10 clips.

Table 4.1 - Responses of All Participants Expressed as a Percentage

Decision	Clip Number									
	1	2	3	4	5	6	7	8	9	10
No action – play on	2	13	9	2	21 <sub>c</sub>	5			18	1
Not enough info <sup>1</sup>	7	6	2		2	2	1		2	
Manage situation	2	8	3	2	10	6		2	16	1
Advantage			5	2	2	1	2		1	1
Penalty to attack	82 <sub>c</sub>	15 <sub>c</sub>	7	31 <sub>c</sub>	15	5	89 <sub>c</sub>	17	11	17
Penalty to defence	5	7	58 <sub>c</sub>	48	15	30	6	12	30 <sub>c</sub>	70 <sub>c</sub>
Free kick			11							
Scrum	2	51	2	10	22	49 <sub>c</sub>	2	55 <sub>c</sub>	18	5
Scrum with turnover			2	6	14	2	1	14	4	5
Level of Confidence in Decision from 1 ( <i>low</i> ) to 5 ( <i>high</i> )										
<i>M</i>	4.1	3.6	3.8	4.3	4.0	3.8	4.6	3.9	3.9	4.2
<i>SD</i>	0.9	1.0	1.1	0.8	0.9	1.1	0.7	1.0	1.0	.09
Kappa statistic	.60*	.14	.20	.14	-.04	.17	.74*	.21	-.01	.41 <sup>AS</sup>
Strength of agreement <sup>1</sup>	Mod	Poor	Poor	Poor	V	Poor	Good	Fair	V	Mod
					Poor				Poor	

The mean level of accuracy for all participants across all clips was 49.6% ( $K = .25$ ). The mean level of confidence in their decisions for all participants across all clips was  $M = 4.0$  ( $SD = 1.0$ ). Significant Kappa statistic indicates a better than chance agreement (significance adjusted by Bonferroni method).

<sub>c</sub> Correct decision.

Strength of agreement<sup>1</sup> as per Landis and Koch (1977)

<sup>AS</sup> Approaching Significance, \*  $p < .05$ .

*Table 4.2 - Percentage of Correct Responses, Agreement and Confidence Scores by Group*

Clip	Referees			Touch-Judges	Assessors	Referee Coaches
	Top-20	21-40	41-65			
1	100.0	87.5	82.6	83.0	70.4	84.6
2	7.1	0.0	18.2	6.4	38.5	7.7
3	64.3	75.0	45.5	65.9	64.0	18.2
4	42.9	25.0	45.5	25.5	37.0	0.0
5	14.3	12.5	26.1	23.9	22.2	8.3
6	78.6	50.0	36.4	44.7	40.7	66.7
7	100.0	71.4	100.0	79.1	92.3	92.3
8	42.9	62.5	52.2	51.1	55.6	76.9
9	28.6	37.5	39.1	28.3	33.3	8.3
10	64.3	50.0	78.3	73.9	70.4	66.7
<i>M</i>	54.3	47.1	52.4	48.2	52.4	43.0
<i>SD</i>	32.9	28.4	26.3	26.7	21.7	37.3

Kappa Statistic of Agreement

<i>M</i>	.39	.30	.35	.29	.29	.34
<i>SD</i>	.32	.20	.27	.19	.21	.26

Level of Confidence in Decision

<i>M</i>	4.3	4.4	4.1	3.9	3.9	3.9
<i>SD</i>	0.8	0.7	0.8	0.9	1.0	1.1

The mean accuracy of all the referees across all 10 clips was  $M = 51.3\%$  ( $SD = 28.5\%$ ). The mean accuracy of all the support groups (touch-judges, assessors and referee coaches) across all 10 clips was  $M = 47.9\%$  ( $SD = 28.6\%$ ).



Investigating the prevalence of SMMs by measuring the extent of shared reasons underpinning decisions revealed perfect coherence when groups displayed perfect accuracy. For example, in clip 1 the top-20 referees achieved 100% accuracy (see Table 4.2) and all chose to penalise the defending players for encroaching offside. Similarly in clip 7, both the top-20 and lower-panel referees revealed maximum accuracy with 100% agreement, since all the participants awarded a penalty to the attacking team for the defender's failure to roll away. In addition, across all 10 clips when officials were accurate, the top-20 showed a considerably higher level of coherence in their reasons underlying decisions ( $M = 93\%$ ) when compared to all other groups (mid-panel,  $M = 86\%$ ; lower-panel,  $M = 82\%$ ; touch-judges,  $M = 80\%$ ; assessors,  $M = 87\%$ ; referee coaches,  $M = 80\%$ ).

#### *4.43 Areas of Concern in Refereeing the Tackle*

Surprisingly, given the level of officials examined, 2 of the 10 clips revealed extremely low accuracy scores (clip 2,  $M = 15\%$ ; clip 5,  $M = 21\%$ ). Furthermore, in an additional three clips, participants failed to achieve 50% accuracy (clip 4,  $M = 31\%$ ; clip 6,  $M = 49\%$ ; clip 9,  $M = 30\%$ ). Moreover, when the levels of agreement are considered, clip 5 and clip 9 revealed a negative kappa statistic (clip 5,  $K = -.04$ ; clip 9,  $K = -.01$ ); a result in fact lower than the level that would be predicted by chance alone. Interestingly, for clip 5 there was no drop in confidence levels ( $M = 4.0$ ) across all the participants. In fact, they were nearly as confident in this decision as they were for the first clip ( $M = 4.1$ ) where 82% of them made an accurate response.

Further exploration into the coherence of the reasoning underpinning decisions revealed the greatest discrepancy in clips 2 and 9. In clip 2, where only 15% achieved the correct decision, 68% awarded this penalty for support players arriving off their feet, 19% for the tackler not rolling away, and 14% for offside (all legitimate rulings within the laws). Similarly in clip 9 the participants were divided, with 51% awarding the penalty for not releasing the ball and 49% for the ball carrier's support arriving off their feet.

From an applied perspective, clip 4, as well as showing relatively low levels of accuracy (31% awarding a penalty to the attacking team) also resulted in 48% of the participants awarding a penalty to the defensive team and 45% awarding possession to the attacking team, either through awarding a scrum, playing advantage or choosing to play on. Thus the participants were almost equally split on which team should benefit from the decision, which would clearly have a profound effect on the game. Although the two experts had initially disagreed on this clip, they were both in agreement that the attacking team should benefit from the play.

Similarly in clip 7, while producing high levels of accuracy ( $M = 89\%$ ) 13% of participants believed that the clip contained an offence worthy of a yellow-card, a procedure used to sanction a player and send him off the field of play for 10 minutes. Once again the levels of confidence in the accuracy of the participant's decisions ( $M = 4.6$ ) did not reflect this DM discrepancy.

#### *4.44 The Fidelity of the Video Recordings and the Naturalistic Paradigm*

The participants' feedback suggests that the NDM procedure used in this investigation was acceptable for all the groups examined. Only 26 of the 1,320 participant responses (i.e., 132 participants assessing 10 situations) were reported as holding insufficient information to make a decision, while the mean confidence level for all participants across all clips was  $M = 4.0$  ( $SD = 1.0$ ) out of a maximum of 5.

In terms of the ecological validity of the procedure only 14% of the participants believed that the quality of the video and camera angle needed improving in at least one of the clips, although no consistent pattern emerged as to which clips needed enhancement. Also, 10% suggested that more information on the game such as scoreline and knowledge of previous plays would have made the decision easier. Only 5 of the 132 participants made comments on the influence of the referee on the screen. However, all the participants felt the test to be a fair evaluation of referee DM prowess and most pertinently for the present investigation, there was no relationship between negative feedback on the information presented on the screen with the levels of accuracy or agreement shown.

### **4.5 Discussion & Conclusions**

#### *4.51 Analyses of all Participants*

The primary aim of this phase was to assess the accuracy, agreement and coherence of England's best RFU referees, touch-judges, assessors and referee coaches. The

mean levels of accuracy and agreement revealed poor DM performance. Despite selecting difficult DM scenarios all 10 clips were judged by experts to be representative of actual decisions required on the field of play, which occur on average 11 times per game. Since these RFU officials averaged only 50% accuracy, this represents approximately 5 or 6 wrong decisions per game. Clearly, the ramifications on the game may be significant. Moreover, it is of even greater concern that the participants' level of confidence in their decisions rarely decreased even when their decisions became more discordant. In other words, although these top officials made both inaccurate and widespread decisions, they were all as individuals equally confident in the accuracy of their DM.

#### *4.52 Efficacy of SMMs to Test Officials' DM*

As suggested earlier in this chapter SMMs did appear to help accurate DM since when a high percentage of officials were accurate their shared understanding as indicated by the same reasoning was also high. Equally, when the number of accurate responses was low the reasoning underpinning those decisions was even lower. In addition, since the top-20 referees indicated greater coherence in the reasons underpinning their accurate decisions, it seems fair to conclude that their mental models had more similarities. This supports the ecological and criterion validity of the methods used.

The critical emphasis in the development of SMMs relies on understanding the reason for differences in decisions. The simplest explanation may be that the different

decisions are a reflection of the participants' ability to identify the cues pertinent to making an informed decision. In fact, as outlined by Mortimer and Collins (1997) it may be that the individual participant had a particular scaling value for the pertinent cues, using the terms *criteria* (the recognition of relevant cues) and *weighting* (the relative value of each of the criteria in reaching the decision). Thus, in the rugby union tackle situation one referee may rate the tackler's inability to roll away as the most important criterion above the ball carrier's decision to hold on to the ball until support arrives. This would result in awarding a penalty to the attacking team. However, if another referee weighted the ball carrier's obligation to immediately pass, place, or release the ball as more important then this referee would be more likely to award a penalty to the defending team. This may explain the poor coherence levels in clips 2 and 9. Accordingly, applying a hierarchical weighting scale where elements of the decision are prioritised may be one method of improving DM in such highly time pressured environments (Annett, 1997; Rasmussen, 1985).

#### *4.53 Analyses by Group*

Some inter-group differences were also apparent. For example, collectively the referees were marginally better than the support groups. However, the mid-panel referees' performance was worse than both the touch-judges and the assessors, yet they were the most confident in their decisions. This may suggest that the mid-panel referees achieved this level of ranking because of their greater confidence levels rather than through more accurate DM. A study by Franks, Elliott and Johnson (1985) would seem to support this idea. Their investigation asked expert and novice

gymnasts to view paired handspring performances to identify if there were differences between the two and to state where these differences occurred. Results showed that the experts were no more accurate but were simply more confident in their decisions. However, in the present case the super-elite top-20 group, which included several international referees, showed more “realistic” confidence scores since these levels more accurately represented their levels of accuracy and coherence.

Most alarmingly the referee coaches revealed the lowest levels of accuracy. In fact, they were worse than the referees to whom they are required to offer guidance. Since most of these individuals are ex-referees who had not performed in many years, this is perhaps not surprising since the speed of the game is now much quicker (Campsall, 2002) and inevitably interpretations have similarly evolved to meet the new demands of the professional game. Nevertheless, this has enormous implications for the development of elite referees. If the referee coaches, the individuals responsible for teaching referees, are offering erroneous or disparate advice on this critical area of law-application, the current levels of inaccurate and incoherent DM may remain.

Before concluding, it is important to consider any methodological limitations that may have contributed to the findings. First, it is possible that some of the officials may have seen some of the test incidents before as they may have been broadcast on television or indeed the participants may have been involved in officiating the games. However, a referee would typically officiate in at least 25 games per season, each

containing in the region of 120 tackles, which would total about 3,000 of these types of situations. In this intervention it is questionable whether or not referees would have been able to remember each incident. Nevertheless, more control should be taken to prevent this in future investigations.

Secondly, another possible limitation of this phase is the small number of test clips that were used to assess referee performance. Furthermore these *difficult* clips were not necessarily representative of the most common tackles that are likely to be encountered. Therefore, future studies should investigate a wider variety of scenarios in order to explore the levels of accuracy and coherence that are required to referee at the top level. More importantly, to help ensure that game outcomes are not adversely influenced by poor referee decisions, interventions should provide an expert's detailed interpretations, focusing on the types of tackles that create problems in order to produce more coherent referee DM.

#### *4.54 Applied Implications*

From an applied perspective, this video-based NDM approach offered a means of identifying areas of concern (cf., Abernethy, 1996). For example, the findings revealed inconsistent use of yellow-cards (and the subsequent loss of a player for 10 minutes) and revealed decisions with penalties awarded in opposite directions.

It is surprising that two clips (5 and 9) revealed levels of agreement lower than that which would be expected by chance (as reflected by the negative Kappa scores in

Table 4.1). Thus, for these two specific cases taken from premier league games, the decisions made by England's best RFU referees (which included two international referees ranked in the world top-20), touch-judges, assessors, and coaches appear to offer decisions that could justly be described as random. It appears that with respect to the application of law 15 England's top officials provided very unpredictable decisions. Clearly, the influence of such poor coherence on the game can be substantial, with players having to adjust their play week by week to fit in with the individual foibles of each particular referee. Perhaps this is acceptable, although there is currently no data indicating the level of consistency that is acceptable in any sports setting. Nevertheless, the views expressed by premier league coaches at the time, the main consumers in this case, were clear. They wanted a lot more consistency and see the development of greater coherence in the management of law 15 as the most critical factor for the improvement of RFU refereeing (Bunting, 1998; Melrose, 1998). Finally, although only very few clips generated this disappointingly low level of agreement, the impact of such extremes on player trust and the respect held for officials may have wider implications. In short, one inaccurate decision especially at the wrong time could change the tenor of the whole match.

In addition to highlighting particular types of tackles that created problems, this test also identified the groups of officials who were less accurate. The referee coaches' poor performance in particular, strongly suggested the need for some form of SMM training; in fact this was subsequently built into the overall support programme.



Developing the declarative knowledge of the task, the key concepts and their inter-relationship, by exposing the expert's reasons underpinning their decisions might be an appropriate way to improve their understanding of the tackle (Stout et al., 1996). This phase also highlighted the need for future research to examine the efficacy of such techniques for sports officials in light of the growing literature in NDM (Cannon-Bowers et al., 1996).

#### *4.55 A Naturalistic Approach to Referee Decision-Making*

The findings of this phase supported McLennan and Omodei's (1996) conclusions that "own-point-of-view" video scenarios, in this case closely representing the match day referee's perspective, can effectively be used for investigating referee DM through a NDM methodology. All participants indicated that this approach represented a fair test of their refereeing prowess, while suggestions for refinement were relatively minor. Furthermore, all the participants showed high confidence levels and only a very low percentage were unable to offer decisions due to insufficient information.

Despite this support for the NDM framework, feedback suggested that several other factors might need to be refined in order to make the test and subsequent training systems as real as possible. For example, some participants felt that knowledge of the flow of the game may be beneficial, with comments such as "it didn't allow me to get a feel for the atmosphere" and "it would have been helpful to have seen previous

plays in the game.” However, these are factors that may be more representative of the art rather than the science of refereeing (i.e., the judgment of context rather than pure law-application). While it may be argued that context forms a critical part of “mastery of the laws” (see Anshel, 1995; Anshel & Webb, 1991) it seems sensible that before developing such advanced skills like contextual judgment, officials develop coherence in pure law application, which provides the critical foundation upon which to develop more advanced skills. Without such, officials may become even more discordant as contextual factors are added. So, in the absence of contextual factors such as the emotion of the players and the tenor and flow of the game, the present assessment provided a clear, unambiguous test that required a comparatively unequivocal application of the law.

In keeping with the literature, this phase supported the contention that researchers need to look at the reasons underlying decisions as well as the actual decisions made. Thus, to address the problems highlighted here, the next phase of the support designed a training package using these types of “contentious,” yet realistic refereeing scenarios to expose an expert’s mental model. This was intended to accelerate the process of amassing experience (Stokes et al., 1997) and advance the development of SMMs so that referees decisions are not esoteric but rather based on an accurate and coherent understanding of law.

# Chapter 5. Training Accuracy and

## Coherence in Law Application

### 5.1 Preamble

Given the poor levels of accuracy, agreement and coherence for referees, found in Chapter Four, a systematic and scientifically validated approach to training the application of law was required. This would give referees the opportunity to conduct cognitively based training outside of the performance environment and thus shortcut the road to expertise and address the request from premier league coaches to improve referees' consistency in the application of the laws (Melrose, 1998).

### 5.2 Introduction

At present, referees appear to rely solely on experience to develop expertise. Since a large body of research suggests that expertise requires 10, 000 hours of deliberate practice to master, both in the sporting environment (Helsen, Hodges, Van Winckel & Starkes, 2000; Helsen, Starkes & Hodges, 1998; Starkes, 2000) and in more cognitively orientated activities such as chess (Simon & Chase 1973), this purely experiential approach would equate to about 7,000 rugby union games; the equivalent of refereeing one game per day for 35 years, assuming the season is 200 days long (Glaser, 1984). Moreover, research by Williams and Davids (1995) suggests that mere experience will not necessarily lead to expertise. Refereeing itself may not provide a sufficient number of challenging scenarios and in close succession to develop expert performance (cf. Means, et al., 1993; Starkes & Lindley, 1994). In

short, purely “learning on the job” appears to be a questionable strategy and some form of specific training in this crucial component of the referee’s art is essential. The naturalistic paradigm, offers a valuable approach (Cannon-Bowers et al., 1996).

### *5.21 Developing Decision Making: Applying NDM to Refereeing*

As described in Chapter Three, in an effort to explore the shared nature of knowledge structures, researchers have examined the extent to which mental models are common among team members (Webber, Chen, Payne, Marsh & Zaccaro, 2000). Cannon-Bowers, et al., (1990) suggest that such SMMs are the key to team DM, allowing implicit coordination through a shared understanding of (a) the problem definition, (b) the plans and strategies for solving the problem, (c) the interpretation of cues and information, and (d) the roles and responsibilities of the team members (Orasanu, 1990). Thus, SMMs seem to assist both interdependent DM, which is essential to a referee performing in a team of three (that is when accompanied by two touch-judges) and independent DM, assisting different referees to officiate in the same way, week after week. In addition, they provide the key to measuring refereeing performance success, as the reasons that underpin each decision will reflect the SMM (see Langan-Fox et al., 2000). When both decision and reasoning are correct the referee has a complete mental model of the situation and can truly be described as “accurate”. When different referees share these two-levels of accuracy they can be described as “coherent.” Consequently, players will experience more consistent and understandable decisions, increasing both player and spectator satisfaction.

### *5.22 Training Accuracy and Coherence with Scenarios*

Effective teams need a variety of complex mental models to form an understanding and generate predictions about likely events (Salas, Cannon-Bowers & Johnston, 1997). Accordingly, presenting “typical,” filmed scenarios with cues and patterns that characterise the domain may be an attractive alternative to “on-the-job” learning (Means et al., 1993; Cohen, Freeman & Thompson, 1997).

Using scenarios offers the flexibility to provide clear and timely feedback rather than waiting until some time afterwards, which may be long after the incident under scrutiny has occurred (Means et al., 1993). In addition, a variety of scenarios can be provided that might otherwise take years to encounter. Practice problems can also be tailored specifically to challenge the individual decision-maker’s current level of performance. Calderwood, Crandall and Klein (1987) suggest that novice decision-makers overlook or are unable to see the important information, thus preventing them from generating accurate and complete mental representations of the situation (Helsen & Pauwells, 1993; Williams, et al., 1994). So, since the experts’ ability to discern the important from the irrelevant information is a key determinant that distinguishes them from novices (Abernethy & Russell, 1987; Williams, et al., 1999), training scenarios can be designed to account for such process measures (the reasoning) by directing attention to the significant features of a problem (Rouse & Morris, 1986) rather than by using outcome measures alone (Brannick & Prince, 1997). Typically this would be done explicitly by building SMMs of how situations should be refereed, to provide less experienced referees with the knowledge to search for salient cues, how they relate to each other and how they should impose a

weighting to this information in a variety of different situations (see Balke, Hammond, & Meyer, 1973). Rasmussen (1985) termed this weighting scale, *hierarchical task analysis*, suggesting that exposure to this functional hierarchy of subtasks serves to direct attention, reduce mental workload and enhance accuracy.

### *5.23 Purpose of the Study*

The primary aim of this phase was to pilot the use of a video-based coherence training-program designed to accumulate and reinforce referees' SMMs. Following NDM guidelines the aim was to use very specific, realistic, knowledge-rich scenarios, presenting ambiguous information with auditory interference (represented by crowd noise) and demand a time-pressured response (Johnston, Poirier & Smith-Jentsch, 1998). It was hypothesised that by presenting a variety of scenarios with detailed reasoning, provided by a "high-status" expert (see McCullagh, 1986) to reinforce a DM hierarchy (Rasmussen, 1985; Eylon & Reif, 1984), referees would show improvements in both the accuracy and coherence of their decisions from pre to posttest. A final aim was to collect feedback from the referees to obtain more qualitative insights into the efficacy and mechanisms of such training.

## **5.3 Method**

### *5.31 Participants*

RFU referees (n=56) ranging in age from 26 to 51 years and who had officiated on the national panel from 1 to 16 years, volunteered to take part. As a feature of their continuing professional development these referees attended 6-weekly development meetings held at four regions across England. This phase of the research programme

took place at a series of these regional meetings.

A group of 41 referees who attended both the pre and posttests held at two consecutive meetings were assigned to the experimental (coherence training) group. The remaining referees were assigned to either a passive control group ( $n = 7$ ) or an active control group ( $n = 8$ ). The passive control referees were those who were unavailable between the two tests as they were either on holiday between the meetings ( $n = 5$ ), moving house ( $n = 1$ ), or unable to get access to a video recorder ( $n = 1$ ). Based on their national ranking the group of active control participants were asked to watch the same tackles as the experimental group but without the expert's interpretations, on the understanding that they could complete the training after the study. Referees who failed to attend both tests or failed to complete all aspects of the training were eliminated from the study.

Specifically, I was interested in noting the differential effect of training referees performing at different levels. A referee's career progression is determined by improving his or her standing on a national ranking system, made by a group of referee development officers from the periodical evaluations from 37 advisors. This placed referees into 1 of 3 groups, a top-20 group ( $n = 11$ ) who were responsible for refereeing at premier league level, a mid-panel group ranked from 21-47 ( $n = 17$ ) responsible for national leagues 1 and 2, and a lower-panel group ranked from 48-65 ( $n = 13$ ) who officiated at national league level 3 and 4. Each of these groups were further subdivided into two balanced groups with similar numbers and rankings, enabling a reverse baseline test. Although lacking in scientific rigour (as with so

many similar systems) this ranking system represented the basis for reward and recognition and even progression to the pinnacle of international officiating. As such, it held considerable ecological validity and evaluation of the training package with referees performing at different levels.

The control group consisted of 15 referees from the various ranked groupings (top-20,  $n = 3$ ; mid-panel,  $n = 7$ ; lower-panel,  $n = 5$ ), again sub-divided into two balanced groups. An independent t-test between the collective mean ranking of the training groups ( $M = 35.54$ ,  $SD = 16.71$ ) and the control groups ( $M = 35.13$ ,  $SD = 19.34$ ) showed no significant differences,  $t(1, 54) = .077$ ,  $p = .94$ .

### *5.32 Developing the Assessment and Coherence Training Tapes*

Using a Panasonic AGDP800HEG S-VHS camcorder, 12 English RFU premier league games were recorded onto a series of Super VHS videotapes by myself who had worked as a professional camera operator and videotape editor over a period of 12 years. The camcorder was mobile, supported by a monopod for stability, allowing the camera operator to move up and down the sideline of the pitch and stay level with the play. At every break in play (when a penalty, scrum, or lineout was awarded) I moved adjacent to the action, providing an angle similar to that which the match-day referee adopted. This “own-point-of-view” recording has been shown to provide minimum distortion of the complexity and dynamics of naturalistic environments (Omodei, McLennan & Whitford, 1998).

From these tapes, 126 tackle incidents were edited onto a master tape. An



independent expert panel consisting of the three highest ranked referees in England (who were not participants in the study) rated each tackle on ambiguity and quality of information available. Then, all the tackles that yielded a consensus opinion from the three independent experts and were adjudged to exhibit sufficient information and present realistic match situations for refereeing were ordered in terms of ambiguity. The experts rated these remaining 45 most ambiguous tackles on difficulty, providing two balanced groups of 10 tackles (labelled A1 - A10, and Z1 - Z10) for the reverse baseline pre and posttest, and 5 sets of 5 training tackles. To further validate the two groups of tackles as balanced, an independent t-test showed no significant differences between these A and Z tackles on pretest accuracy scores  $t(1, 54) = 1.735, p = >.05$ ).

*5.321 Pre and Posttest Assessment Tapes.* Each of the 20 pre and posttest clips commenced with a voice-over that introduced the two teams competing, indicating which team had possession and the direction in which they were attacking. The tackle incident then began with approximately 5 seconds of “lead-in,” the period confirmed by pilot and previous studies to be necessary to allow the participants to orientate themselves to the scene. After the tackle incident the videotape image froze, presented the title “make your decision now,” and cut to a blank screen after about 5 seconds.

*5.322 The Coherence Training Tape.* This contained the 25 tackles used for training, edited in exactly the same manner only arranged into 5 sets of 5 clips. After each set of five clips the tackles were replayed followed by Ed Morrison (at that time

the number 1 ranked RFU referee) giving a detailed explanation of his interpretation to the camera, concluding with a further rerun of each tackle clip. Consistent throughout his interpretations was the hierarchical theme of (a) getting the tackler to roll away immediately, (b) allowing the ball carrier to release the ball, and (c) ensuring that support players arrive on their feet. Ed Morrison was chosen as the expert to provide the model interpretations, as an active referee whose decisions would reflect the way in which the English game should be officiated. Furthermore, having refereed the world cup final in 1995 and as England's most capped international referee at that time, it was anticipated that his opinions would have credibility with the participants (see McCullagh, 1986).

### *5.33 Instrumentation*

A response sheet (see Appendix Three) was developed to enable participants to quickly and easily signify their decision. The first section of this sheet consisted of a series of six boxes and asked the respondent to tick the appropriate box. This presented options to (a) play on, (b) award a penalty to the defending team, (c) award a penalty to the attacking team, (d) award a scrum to the defending team, (e) award a scrum to the attacking team, and (f) other, with a space to explain this "other" decision. The response sheet then asked participants to explain the reasons behind each decision and indicate on a Likert scale their confidence in the accuracy of each decision, scored from 1 (*low*) to 5 (*high*).

A coherence training booklet was given to each participant in the experimental group, which in addition to having 25 response sheets (one for each of the training

tackles) identical to those for the pre and posttests, included a viewing log to record how often and on which dates they watched the two training tapes (see Appendix Four).” At the end of the booklet a feedback sheet asked the participants to comment on the quality, value and ecological validity of the video training package.

### *5.34 Procedure*

*5.341 Pre and Posttests.* All the participants completed the pretest at their regional development meeting. The posttest was carried out in exactly the same fashion at the next meeting 6 weeks later. I was careful to ensure that all groups were treated in the same fashion, by following a clear set of instructions consistently for each region (see Appendix Five). There were no more than 18 participants watching the assessment clips at any one time, in any region, and as in the previous phase all participants were informed that their own personal responses would remain confidential and that the results would only be presented on a group basis.

After the participants familiarised themselves with the response sheet they adopted a position where they could comfortably see the tackle incidents projected onto a screen via a standard VHS video recorder and data-projector. This presented an image about 6 feet wide and 4.5 feet high. The assessment videotape was then replayed, showing either clip A1 or clip Z1 depending on the group to which the referees were assigned. The video was paused immediately after each clip and participants were asked to make an immediate decision by ticking one of the six boxes. They were explicitly told not to change their decision once made. Both inspection of the response sheets and observation of participants revealed that they

all conformed to these instructions. The participants were then given 3 minutes to write down the reasons for their decision and indicate their confidence scores.

This procedure was conducted by myself and followed consistently for both experimental and control groups in all four regions. With permission from the RFU all the participants conducted no additional, structured DM or video-based training during the whole testing period. This might typically have included group discussions prompted by match-day recordings or the RFU releasing updates on interpretations of law.

*5.342 Experimental Group.* Each experimental group referee was given a copy of the coherence training tape and booklet and asked to watch the tape in the same manner on a set night each week and to record each inspection in the booklet's viewing log. After viewing the first five clips once and completing response sheets, they then watched the model interpretation section and reviewed these same clips as many times as they felt necessary during the week to understand the model interpretation. The viewing logs revealed that all the referees were both reliable in conducting their first viewing on the same night each week and viewed the model interpretation section at least one other additional time before the following week's training. These participants then completed the acceptance sheet and finally after all the training they completed the feedback section, commenting on the efficacy of the training package (see Appendix Six).

*5.343 Control Groups.* To control for any expectancy and potential training effects that might occur merely as a consequence of watching a greater number of tackle incidents, the active control group watched exactly the same tackle clips as the training group but without the model interpretations. They were also asked to re-examine the clips at least one other time later in the week. Examination of their viewing logs revealed that they conformed to this request. All the other control group members were “passive,” simply watching the pre and posttest clips in the same fashion as all the other participants.

### *5.35 Data Analysis*

The primary analysis was quantitative, measuring the accuracy and coherence of participant’s responses. Secondly, the qualitative data from the acceptance sheets and feedback sheets were used to identify trends, providing evidence to show how changes may have occurred.

Quantitatively, the participant’s responses were only considered to be accurate if they provided both the correct decision and the correct reasons underpinning that decision, as deemed by the expert. Thus, if a referee arrived at the correct decision but with incorrect reasoning this was considered to be inaccurate and was grouped with responses that reflected the wrong decision. Referees sharing these two levels of accuracy were considered to be coherent; hence, coherence was the percentage of accurate participants.

The referees' qualitative feedback on the validity and value of the test and the reasons for their acceptance levels were transcribed and analysed by group. In addition to noting the comments, simple evaluations were made using a frequency count of similar responses. A second researcher, unconnected to the study examined the frequency and distribution of these comments and felt that all were valid, credible and reliable, thus reaching 100% agreement with me.

## 5.4 Results

### 5.41 Quantitative Results

Before examining the effects of the intervention it was first necessary to establish the equivalence (or lack thereof) of the two control groups. Accordingly, a 2 x 2 (Group x Time) ANOVA with repeated measures on the second factor, comparing pre and post performance achieved by the active and passive controls. No significant differences were apparent on any of the effects, time:  $F(1,12) = 1.29, p = .28$ ; Time x Group:  $F(1, 12) = .33, p = .58$ ; or group:  $F(1, 12) = .12, p = .78$ . Accordingly, these two were combined as a single control group in all subsequent analyses. These nonsignificant results also meant that any changes in the experimental group's performance were due to the intervention and not just an artefact of viewing the tapes, either through expectancy or just through a greater amount of deliberate practice.

The pre and post intervention accuracy scores achieved by the different groups (presented in Table 5.1) show very high standard deviations, signifying very large variance within each group. Changes in performance were examined by a 4 x 2

(Group x Time) ANOVA, with repeated measures on the second factor. The dependant variable was the percentage accuracy achieved.

Two significant effects were apparent; the main effect of time pre and post the intervention  $F(1, 52) = 5.06, p = .029$  and the group by time interaction  $F(3, 52) = 3.01, p = .038$ . Effect sizes (eta squared) were .071 and .146 respectively, reflecting medium and large values (Clark-Carter, 2001). The main effect of group was nonsignificant  $F(3,52) = 1.26, p = .298$ , effect size = .065, power = .32. Follow up using Scheffé tests suggested that the interaction effect was due to the significant improvement in performance of the lower ranked referees compared to the control group. No other changes achieved significance.

A second 4 x 2 (Group x Time) ANOVA was used to examine changes in confidence scores (see Table 5.1), which revealed a significant group effect  $F(3,52) = 3.34, p = .026$ , effect size = .162. Follow up with the Scheffé test showed that this was due to differences between the top-20 referees and the lower-panel referees. The time effect of confidence across all participants showed a small increase approaching significance  $F(1,52) = 2.94, p = .092$ , effect size = .053, changing from  $M = 3.95\%$  ( $SD = 1.13$ ) to  $M = 4.10\%$  ( $SD = 1.05$ ), whereas the control group's time effect showed a nonsignificant decrease from pre to posttest  $F(1,12) = 1.08 p = .52$ .

Table 5.1 - Percentage Accuracy by Group

	All Training									
	Top-20		Mid-Panel		Lower-Panel		All Controls		Groups	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pre % Accuracy	46.91	11.40	47.06	13.24	39.75	15.24	45.87	13.16	44.70	13.56
Post % Accuracy	50.55	14.81	51.73	13.61	57.18	12.53	41.25	14.44	53.14	13.57
Accuracy Change	3.64		4.67		17.43*		-4.62		8.44	
Pre Confidence	4.05	1.21	4.05	1.04	3.69	1.15	4.17	.92	3.94	1.13
Post Confidence	4.31*	.95	4.07	1.00	4.02	1.15	4.11	1.03	4.12	1.04
Confidence Change	.26		.02		.32		-.06		.18	

$p = < .05$

Finally, to offer a preliminary insight into the mechanisms underlying the changes observed, a post hoc examination of the percentage accuracy increases was conducted on each individual tackle clip. Two tackles showed the most marked improvements from pre to posttests, clip A8 (*pre* = 33%, *post* = 79%) and clip Z2 (*pre* = 35%, *post* = 91%). Both these tackles presented situations whereby in the expert’s interpretation “the ball carrier had become slightly isolated from his support and failed to release the ball.” These data pertaining to the collective understanding



of the situation are considered later in the discussion.

#### *5.42 Referee Qualitative Feedback*

All the participants in the experimental groups reported the process as valuable and worthwhile with comments such as “the video is much better than words” and “it helps me visualise what players should and should not do.” In addition, the referees felt the assessment tackles to be good representations of game situations, with each tackle clip yielding sufficient information to make a decision.

The viewing logs revealed that all the referee groups watched each tackle approximately the same number of times, on average about six times per clip. Nevertheless, there were 16 comments from the lower-panel referees on having to see the tackle more than once, whereas none of the higher ranked referees noted such. For example, one lower-panel referee recognised his initial mistake, explaining “I can see more clearly on the 2<sup>nd</sup>/3<sup>rd</sup>/4<sup>th</sup> viewing,” and another remarked “On the 2<sup>nd</sup> viewing I agree [with the model interpretation], it is clear that the tackler makes little effort.” Seven of these lower-panel referees also commented on gaining a greater understanding of the priorities at the tackle as summarised by one referee who said “I feel I became more aware of my priorities at the tackle after watching the clips.” This improved clarity of the priority system was described by another referee who suggested that “the training package does very well, reinforcing the sequence – (a) did the tackler move, (b) did the tackled player release the ball, and (c) did the next players arrive on their feet.” By comparison, the mid-panel group only made only one such comment while the top-20 referees made no comments at all. When

responding to the training tackles the higher ranked referee groups (top-20 and mid-panel) tended to offer more detailed reasons underpinning their decisions than the lower-panel referees.

## 5.5 Discussion

The findings suggest that this video-based training package was appropriate for developing SMMs in pre-elite referees. Referee comments reinforced the efficacy of using video scenarios to train accurate and coherent DM. Specifically, the lower-panel group who experienced the most improvement felt that their exposure to the model's priority system (cf. hierarchical task analysis - Rasmussen, 1985) was the reason for their enhanced performance.

In contrast, it is perhaps surprising that both the top-20 and mid-panel ranked referees showed only small, nonsignificant improvements in both performance and confidence scores. However, support for the training from the participants was overwhelming and all groups adjudged the exercise to be valuable. Importantly, the referees reported the assessment tackles to be a fair test of referee DM, accurately reflecting decisions that had to be made on the field of play.

Encouragingly, this approach provided a means to identify and train problem areas in refereeing. Subsequently, the RFU employed this type of training as a method to both reinforce the philosophy of refereeing the tackle and initiate new interpretations that regularly descended upon them from the iRB, sometimes midway through a season. As a consequence, the tapes that provided such detailed descriptions of law

interpretations were also used to assist players' and coaches' understanding of the refereeing philosophy. Other similar invasion games such as basketball, soccer and hockey could equally benefit by identifying controversial laws, presenting a variety of referee perspective video clips of these laws together with an expert's detailed decisions and the reasons underpinning them in order to develop shared weighting scales amongst referees. Such systematic training would not only speed up referee development but could also be used to increase player and coach understanding which may lead to reduced controversy. Moreover, coach and player education in the application of the law could potentially reduce the number of infringements and lead to more flowing and attractive games.

Further development of SMMs was facilitated through the increased interactions and discussions which now occur at regular referee meetings, driven by coherence-based video exemplars (cf. Kraiger & Wenzel, 1997). Using video, together with developing a coherent language appeared to be the solution to Rouse and Morris' (1986) concerns of capturing mental models that may be largely pictorial through words alone. As a result of this intervention, increasing employment of the model described in Chapter Two and other support work, referees developed a working language, using phrases such as SMMs, coherence and "priorities in the tackle," to conceptualise and describe the nature of their task. As highlighted earlier, the development of this common vocabulary was an essential component of RFU referee training and could be considered to be an intervention in itself via enhanced communication.

### *5.51 Why Did the Lower Ranked Referees Improve?*

While lower-panel referees were often inaccurate in their initial interpretation of the training tackles many explained that after hearing the expert's interpretation and viewing the clip again the reasoning became more apparent. Thus, it may be that their improved coherence in the posttest is as a result of a richer store of incidents in long-term memory, accessible by retrieval cues that pre-primed these referees into making the appropriate decision (McLennan & Omodei, 1996). Such prior knowledge or pre-priming, which may have occurred as a result of improved cue-utilisation (Stokes et al., 1997), has already been found to affect referees decisions in soccer (Jones et al., 2002).

This explanation seems even more plausible since post hoc analysis revealed that tackles in which the most improvement was made were those in which the referees appeared to be pre-primed towards the likely outcome. Specifically, these were incidents where the ball carrier was "isolated from his support" and became susceptible to illegally holding on to the ball.

### *5.52 Why Didn't the Higher Ranked Referees Improve?*

Given the aims of this phase it was equally important to consider the reasons underlying its failure to increase the coherence of the higher ranked referees. This may have been due to two reasons. First, the higher ranked referees were more explicit in their interpretations than their lower ranked peers, both at the pre and posttest. The top-20 referees tended to offer more alternative and face-valid interpretations for each tackle, although these were often markedly different to the

model answers. In other words, whilst these higher-ranking referees had more complex and developed mental models these were not shared across peers. The positive aspects of the elite group's performance mirrored Orasanu's (1990) findings where high-performance cockpit crews were more explicit and revealed a more complex understanding of situations than low-performance pilots, resulting in quicker and safer decisions in emergencies. It also highlighted the importance of the model interpreter. The advantage of using just one expert is that it is more likely that a consistent message will be presented, which was important for this preliminary intervention. However, future studies may find it valuable to present a consensus opinion from a group of international referees to help ensure that the model answers are representative of the way the game should be refereed.

The second reason may be the expert's diminished influence on this group of referees. The top-20 group were much more equivalent in status to the expert and as such may have been more resistant to change their own mental models. Taking these two points together, the more complex and robust mental models of the higher ranked referees, may explain their nonsignificant improvements.

Finally, the lack of significant differences found between the groups must be considered against the comparatively low power, which given that four groups were examined was conservatively calculated as 0.32. In fact, to reach the generally accepted levels of 0.8 (Cohen, 1988) would have required approximately 40 participants per group. Even though this estimate is a conservative worst-case scenario the possibility that the analysis lacked the power to discriminate between

the groups' performance levels should be borne in mind. This distinct possibility notwithstanding however, the improvements generated by the intervention were clear.

### *5.53 Why Only 50% Accuracy?*

While it may be surprising that none of the referee groups achieved even 60% accuracy, one must consider the particular nuances of the sport that was scrutinised in this phase. As discussed in Chapter One, recent referee DM research has explored more "matter of fact" scenarios such as the offside decision in soccer, asking merely whether the player was offside or not? At the next level, Plessner and Betsch (2001) and Jones et al. (2002) considered "matter of opinion" decisions, asking soccer referees to judge whether a foul was committed and if so, by whom. In contrast, however, this phase of the support programme explored a third and hereto unconsidered level of complexity. Refereeing the tackle in more open sports like rugby union presents a unique situation where multiple, complex and dynamic decisions are required, as there are timing elements, overlapping elements, and interactive elements (see Ackford, 2003). In essence, the degrees of freedom in this situation are so great that the level of accuracy demonstrated here might be appropriate.

Furthermore, a rugby union referee is much more than a mere regulator of the law. A feature of refereeing this sport is the notion of advantage. For example, the referee would recognise that an offence has occurred but may choose to ignore it if no advantage has been gained or simply "manage it," perhaps through communicating

to the players to balance the trade-off between game flow and control. Indeed, it is the referee's ability to allow the game to flow but also maintain the control of the players (termed contextual judgment – see Chapter Two) that is crucial. Furthermore, the increased degrees of freedom in rugby union may make the contextual factors even more crucial. This may in part explain why the higher ranked referees were unable to improve upon their initial performance, as contextual judgment at premier league level may well supersede law application as the more critical factor. This area required further investigation but the importance of contextualising results against the specific challenges inherent in the game are well evidenced by this situation.

Nevertheless this chapter, together with work presented in the previous chapter and subsequent investigations, establishes the standard to be as low as just over 50% accuracy, despite these samples including several international and ex-international referees. Interestingly, when 12 of England's elite rugby *league* referees were shown clips of similar tackle situations from their sport (one which presents slightly fewer degrees of freedom or reasons for awarding a penalty than the rugby union tackle) they achieved only 63% accuracy by the same measures employed here (Mascarenhas, 2004).

It must also be recognised that the results here present an extreme score as the referees were only considered to be accurate if they achieved the correct decision and the correct reasoning. Although examining the decision alone may give results more similar to other sports (typified by yes/no and judgment decisions), if they are not accurate on at least these two levels in rugby union one cannot assume any SMM to

be apparent. In fact, since the tackles that proved to be most trainable were ones in which the referees appeared to be pre-primed into their decisions, this seemed to suggest there to be a third level of coherence. Referees should not only award penalties to the correct team and for the same reason but they should also have the same understanding of that situation, which may help to pre-prime them into anticipating the event before it occurs. This has strong theoretical implications for the development of SMMs and remained a consideration for future research.

Another reason for such apparently low scores may be attributed to the tackles that were specifically chosen as a range of difficult situations that regularly occur in rugby union. Nevertheless, whatever the reasons for the levels of coherence found in this phase, it highlighted an intervention that is capable of developing pre-elite referees into line with the elite.

Finally, it is interesting to note, the anecdotal evidence from advisors, coaches and players suggesting that such improvements have transferred to their application of law 15 on the field of play. This may not only be as a result of this intervention *per se* but as a corollary of such work, other aspects of the support programme and the increased interest and discussions that subsequently took place.



## Chapter 6. Referee Decision-Making:

### It's Not What You Give It's The Way That You Give It!

#### 6.1 Preamble

Despite very positive results and feedback on the law-application training, it was becoming increasingly clear that contextual judgment was perhaps more crucial at the top level. If elite referees were only achieving approximately 50% accuracy, then contextual judgment must have a significant impact on their decisions. Consequently, and given the increasing demand for advanced training for full-time referees it became necessary to begin to explore the final cornerstone - contextual judgment. Firstly it was imperative to identify the important contextual factors and secondly explore the impact of these on referee behaviour.

#### 6.2 Introduction

As previous chapters have demonstrated, a referee's knowledge and feel for a game is crucial in applying the spirit of the law, rather than simple adherence to a strict application of the letter of the law (Askins, Carter & Wood, 1981). Since games vary markedly, referees often alter their style of refereeing to suit the particular nuances of the game (Grunska, 1999); a feature identified in Chapter Two. In fact, for certain more open invasion games this judgment may take precedence over a strict and accurate application of the law on isolated incidents (Mascarenhas, Collins, Mortimer & Morris, 2005; Stern, 2002). Such apparent "malleability" in the law may explain at least in part, why previous studies in referee DM (e.g., Nevill et al., 2002) have shown top-class referees to be less than 100% consistent in their application of the

law on a series of snapshot incidents.

The relative importance of such contextual factors is based on the referee's appreciation of the tenor of the game and results in a weighting scale that can supersede their application of the law. Such understanding is particularly crucial in rugby union as it has the potential for open and flowing passages of play. In much the same way as a police officer attempts to prevent crime, the referee in such open sports has the same responsibility, encouraging players to play within the laws so that flow (and spectacle) may be maintained. Thus a referee may penalise a minor infringement, or allow play to develop and manage the situation, or penalise but be particularly expressive in the condemnation of the player involved to prevent the action from reoccurring (Grunska, 1999). These latter two options represent good practice as they assist in developing flow, despite the many calls from the media for referees to provide a consistent and consequently unfeeling application of the law. Accordingly, the weighting that referees apply to their DM is crucial and it is important to understand the factors that affect this weighting. Furthermore, the link between the decision and the way in which the situation is managed, such as how the referee convinces the players of the appropriateness of that decision or how firmly the referee chastises player behaviour is equally important (Askins, 1987). However, to date no empirical research has explored which factors referees use to judge the context of a game, how they manage that context and how universal these are across different levels of the game.

Despite such scant literature examining the DM of sports officials, there is a growing body of literature that focuses on the knowledge underpinning DM in naturalistic environments. For example as described previously, Stout et al. (1996) suggested that knowledge structures, such as those that provide referees with an accurate understanding of the game, are organised into mental models. Mental models enable accurate perceptions of elements in the environment, comprehension of their meaning and prediction of their future status (Endsley, 1988). In much the same way as viewing knowledge as either declarative or procedural, mental models also contain both declarative and procedural information. Converse & Kahler, 1992 suggested that a mental model comprises a declarative mental model (DMM) providing knowledge of “what and why” and containing information about the concepts in the domain and relationships between them, and a procedural mental model (PMM) providing knowledge of how to do something, storing information about the steps that must be taken in order to complete the task or goal (see Stout et al., 1996). As such, to get a complete picture of the mental model underpinning referees’ decisions it is important to access both their DMM, for information on what the decision is based upon and their PMM to understand how they have chosen to manage the situation.

Although not directly examined, there is some empirical evidence for the employment of these two types of knowledge in refereeing. For example, Jones et al. (2002) found that while soccer referees’ prior knowledge of a team’s aggressive tendencies did not

affect their propensity to award more decisions against them, they did receive more yellow and red cards. Jones et al. suggested that this clear signal of intent was to prevent any anticipated future escalations of aggressive behaviour. In fact, such warnings provided by the PMM, often termed “preventative refereeing” are crucial for top-flight referees (Grunska, 1999). Furthermore, such evidence is not limited to empirical research, as national governing bodies also suggest the need for referees to maintain a “balance between game control and game flow, having a feeling for what the participants are trying to do and calling what is right for the game” (FIBA, 2004). Surprisingly, despite this need for player and game empathy the iRB’s assessment system (used to identify and advance good referees) is a series of tick-boxes, demanding a robotic and unfeeling application of the law and providing little opportunity for feedback on what is considered by many to represent the basis of higher order performance; the art of refereeing.

In summary, there is considerable anecdotal and empirical evidence suggesting that contextual judgment is important, highlighting the need for referees to develop an appreciation for the context of the game, although there have been no reported attempts to measure such skills. Accordingly, the purpose of this investigation was first to identify the factors that moderate rugby union referee DM away from a strict application of the law and second, to examine how they use their DMM and PMM to control high-pressure games. Thus a two-phase investigation was adopted to examine the two research questions. Once again the tackle was the focus of

examination as it occurs frequently, is affected by a number of laws, is very challenging to apply, and also has the potential to create flashpoints in the game.

## 6.3 Phase One Method

### *6.31 Participants*

I was interested in finding out the extent to which referees felt that contextual factors exist at different levels of the game. Accordingly, both high potential referees earmarked for fast tracking to higher duties (currently responsible for national leagues 2 and 3) and also elite referees at the top of their game (responsible for both international and English premier league matches) were sampled.

These two independent focus groups volunteered to take part in the study as part of their ongoing professional development. The first group included six, high potential male referees, ranging in age from 23 to 39 years ( $M = 33.17$ ) with a mean of 7.2 years of experience in refereeing and currently performing at national league level. The second group consisted of four, full-time male referees ranging in age from 40 to 48 ( $M = 42.3$ ), averaging over 20 years experience in refereeing and all responsible for premier league and international games. Both groups were informed that their individual responses would remain confidential and that none of their comments would be reported to the RFU management team.

### *6.32 Procedures*

Participants were shown a series of video clips of rugby union tackle situations and asked to write down their decisions for each clip. The following questions were subsequently posed to the group:

1. What might occur around that event (either before or afterwards) that could alter your decision?
2. Might you referee this situation differently in different games or for different individuals? If so, why and in what ways would you modify your DM or behaviour?

In accordance with the nominal group technique (Delbecq & VandeVen, 1971) participants were given a few minutes to think about their responses to these questions and write them down. Then in turn, participants read out factors from their list (without judgment from the group) and I recorded them onto a master list, whilst recording all discussions to tape, for later analyses. The focus group schedules detailing the facilitator's notes, instructions to participants and the focus group response sheets can be found in Appendix Seven.

Specifically focusing on those contextual factors that occur during (not before or after) the game, each group was asked if any items on the list could be collapsed together. I acted as a facilitator, encouraging the groups to discuss each item and to develop a coherent understanding of each item by probing with open-ended

questions. I was careful not to get in the way of participants expressing their opinions, experiences and suggestions, and encouraged all members to contribute. The relationship I had developed with the participants allowed them all to contribute without fear of condemnation. This procedure was conducted consistently for both groups.

### *6.33 Data Confirmation*

To ensure the accuracy of the analysis, the master list of contextual factors together with their allocated themes were posted to participants for member checking (Patton, 2002). They were specifically asked whether the reported items and themes accurately represented what was said and if not, how should the report be modified?

All participants unanimously agreed that the data were an accurate representation of what they said, with only one comment on modifying the allocation to themes. This referee questioned whether the time in the game and scoreline should be collapsed together, as referees may alter their style of refereeing when the scoreline is close towards the end of the game. Though not acknowledged in phase one, this consideration formed part of the analysis for phase two.

## **6.4 Phase One Results**

Figure 6.1 presents the master list of contextual factors generated by both groups of referees and Table 6.1 lists the ratings of importance.

Figure 6.1 - In-Game Contextual Factors Identified by Groups of High Potential and Full-Time Referees





Table 6.1 - Referees “Contextual Factor” Mean Importance Ratings (out of 10)

	High Potential Referees (n = 6)		Full-Time Referees (n = 4)		All sample (n = 10)	
	Mean	SD	Mean	SD	Mean	SD
Temper of the game	8.00	1.10	7.00	1.15	7.60	1.17
Players’ respect/rapport	7.17	1.72	7.00	1.83	7.10	1.66
Position on the pitch	5.67	2.25	6.75	2.06	6.10	2.13
Scoreline	5.50	1.97	6.50	1.29	5.90	1.73
Time in the game	5.00	2.45	6.50	1.73	5.60	2.22
Refereeing the outcome	4.20	3.49	6.00	1.83	5.00	2.87
Conditions (weather/pitch etc)	4.83	1.94	5.00	2.58	4.90	2.08

Both groups rated the temper of the game, players’ respect/rapport and the position on the pitch as the most important contextual factors. All reported that these factors would influence their behaviour, either as a direct influence on the decision to penalise or through modifying how “hard” they were in their interactions with players. The only noticeable discrepancy between the two groups was in “refereeing the

outcome,” where full-time referees showed a mean rating of 6 out of 10 and the high-potential referees rated 4.2 out of 10.

## 6.5 Phase One Discussion

Post-hoc checking of the ratings between both the referee groups showed a substantial level of agreement in both the constructs and their rating. As such, within the limitations of using this small (but deliberately specialised) sample, a clear set of contextual factors emerged. “Refereeing the outcome” was the only exception to the between group consistencies. Analysis of the focus group tape-recordings revealed that high potential referees felt their “refereeing the outcome” scores were largely based on the coach and advisor’s perceptions of the game, whereas the full-time referees interpreted “refereeing the outcome” to be their management of game flow and control, particularly in respect of foul play. For example, one full-time referee offered the example,

“I’ve seen that, it’s a penalty, hold on, hold on, it’s okay the ball is out – play on. You put your guns in the holster but you’ve recognised it and log it for future comment. It’s like a suspended sentence.”

It is perhaps not surprising that the high potential referees were more conscious of the individuals responsible for determining their future promotion. However, this influence was extraneous to the purposes of the investigation as it refers to a contextual factor that occurs after the game and therefore while noted for future attention in support work it was eliminated from further consideration. Following

phase one, I sought to verify the real impact of these contextual factors in game situations, limiting this further examination to the full-time referees in order to glean the most accurate, high level representation for subsequent modelling by trainees.

## 6.6 Phase Two Method

### *6.61 Participants*

Since the full-time referees seemed to provide more game focused contextual factors and this is the group that referees aspire to be like, three of the referees who took part in phase one (two of whom refereed in the last eight world cup final games in 2003) completed this phase of the investigation. This comprised referee A (age = 40, experience = 23 yrs), referee B (age = 41, experience = 24 yrs), and referee C (age = 40, experience = 12 yrs).

### *6.62 Pilot Study*

Based on discussions with participants a traffic light scale was developed for rating the “illegality” of each incident (see adapted scale in Table 6.2) and the degree of “hardness” applied to the players involved. A pilot study was conducted to validate the methods, familiarise participants with the procedure and to check the validity of the response sheets.

As described in many NDM texts, the expert fire-fighter, military commander or in this case the referee does not set out to make a series of decisions but rather sets out

to either control a fire, win a battle or control a game (Means et al., 1993). As such, there was a need to look at decisions in context, as DM involves a whole series of actions that affect the environment in ways that influence subsequent decisions and so on (Cohen, 1993). Therefore, a detailed case study approach was adopted, using games that each participant had refereed approximately 3 days prior (Bloom, 1953).

Participants watched their own game-tape, pausing the tape after each tackle, describing what had happened leading up to and including each incident and rating how illegal each tackle was on a sliding scale ranging from green (1), to amber (2, 3 or 4), or red (5, 6 or 7) (where green = completely legal; amber = bordering on illegal; and red = illegal and therefore merited a penalty). For each tackle, they also graded their level of verbal and non-verbal response as either, none, soft, medium or hard; and provided reasons for this level of response. Their description of events and reasons for their adopted level of response were videotaped for analysis.

In the pilot work, the referees found the scale for decision illegality (1-7) offered insufficient discrimination, so a 10-point scale was adopted for the main study (see Table 6.2 and Appendix Eight). In addition, after reflecting upon the data gathered in the pilot study it emerged that using more challenging games would increase the clarity of contextual factors arising and that participants would need to exhibit a diverse range of conflict management skills, often requiring the use of harder verbal and non-verbal communication.

### 6.63 Procedure

Following advice from the pilot study each participant selected a challenging game for analysis. I felt that mismatches between the severity of illegality of tackles and the severity of the subsequent management would indicate that the context of the game had changed. The game tape was recorded from broadcast television pictures and included an in-game microphone recording to enable the referee to hear his own comments made to players and his team of officials. As in the pilot study, participants were asked to think-aloud following each tackle, during three 15-minute periods of the game; the first and last 15 minutes and another period that the referee deemed to be the most challenging period. This enabled analysis of the changes in context at different points during the game.

*Table 6.2 - "Illegality of Decision" Rating Sheet*

GREEN	AMBER			RED			Y-CARD	O-Card	R-CARD
1	2	3	4	5	6	7	8	9	10

Y-Card = yellow card (sanction used to caution players, requiring them to sit in the sin-bin for 10 minutes)

O-Card = orange card (although this card does not exist, participants felt this was appropriate for quantifying a decision that fell between a yellow and a red card)

R-Card = red card (sanction used to eject a player from the game)

Where cognition is very conscious (albeit quick) such as in refereeing, think-aloud, verbalisation data may be appropriate (Abernethy, Thomas & Thomas, 1993). However, some research has criticised the accuracy of think-aloud tasks, suggesting that participants may provide a rational explanation of what they would expect others to say rather than what they actually experienced. Nevertheless, much of this criticism arose through explorations of movement execution tasks (e.g., Williams & Davids, 1995) rather than the more cognitively dominated task under investigation here. As highlighted by Nisbett and DeCamp Wilson (1977) the *nature* of the verbal report task is crucial to the reliability of reporting. Importantly, referees regularly justify their decisions on the field at the time of execution and are also routinely required to provide the rationale for their decisions to assessors, coaches and the media after games. Consequently, participants were accustomed to verbalising their thoughts and expressed complete comfort with both the procedure and the outcomes of the study.

#### *6.64 Data Analysis*

All of the referees' comments were transcribed verbatim for analysis and consensus validation. Two researchers (myself and another colleague) independently studied the transcripts, tagging each tackle with the referee's underpinning reasons. In cases where the two researchers found discrepancies in their interpretation of the text we collectively reviewed the video of the referee describing the reasons for his decisions and his subsequent response level. Discrepancies were found in only 7 of 251 instances, representing a 97% agreement. In all seven of these cases we gained

unanimous agreement after reviewing the videotape.

The game tapes were analysed alongside the referees' responses and the position of each tackle was recorded onto a scaled grid of the pitch. This provided a positional classification of tackles, first those in the attacking or defending half and then those that were deemed to be kickable or not kickable. The referees collectively agreed that kickable tackles were all those taking place in the opposition's half, except those occurring both within 5m of the sideline and 10m of the halfway line (official line markings on an RFU field of play) forcing an angle too acute and over too great a distance to provide a realistic attempt on goal. The first categorisation was used to check for differences related to attack or defence, while the second was related to the extent to which a penalty, if awarded, would offer a scoring opportunity (realised in rugby by kicking the ball between the posts).

## 6.7 Phase Two Results

The coefficient of correlation between the illegality of the tackle and the corresponding hardness of the referee's response was  $r = .724, p < .01$ ;  $r = .785, p < .01$ ;  $r = .763, p < .01$  for referees A, B and C respectively, with the aggregate correlation  $r = .712, p < .01$ . This good level of agreement suggests that referees' severity of management matches closely to the illegality of the tackle.

Analysis of when mismatches occurred, in particular when the "hardness" of

management increased disproportionately to the illegality of the incident, could reflect a change in the context of the game; accordingly, such cases were systematically examined in relation to each of the constructs identified in phase one (i.e., the “temper of the game,” “level of player respect/rapport,” “scoreline,” “position on pitch,” and “time left in the game”). In addition, these same constructs were examined in all the instances when verbal and non-verbal management increased to either medium or hard responses to check for trends.

### *6.71 Mismatches in Severity of Illegality and Management*

Referee A had 9 mismatches in severity from 98 tackles (i.e., management was “medium” or “hard” yet the play was graded between 1 and 4 and thus legal). Two were deemed to be due to the rising temper of the game, described as an “increased contest for the ball.” For example after one tackle he said, “I’ve got really close in because it was all a contest.... Players’ awareness has gone up so I have to also. It’s more critical now.” The other seven were preventative; two of which were due to the “position on the pitch” and “scoreline,” where referee A commented:

He’s on his feet and there’s a legal turnover, which for this game was quite crucial. Greens were banging up to get within drop goal area and now they’ve turned the ball over. . . . My response goes up to medium, because I’m actually explaining what’s happened, because I think it’s crucial. It’s crucial to be accurate in the tackle and it’s also crucial in relation to this game, the position on the pitch. It’s 13-12.

Referee B had no instances (from 72 tackles) where his game management increased disproportionately and referee C had three such instances from a total of 78 tackles.



All three of these were described as preventative. The first was to prevent a flashpoint in the game, where he commented:

There's red players coming off the sides and I'm talking them out of it. . . . A red player has come straight through the middle of the maul and got the ball. The blue player complained, and I've said, "he's come straight through the middle. I'm happy." . . . My response was medium, because I need to make sure that blue are aware that I'm happy with what's gone on to *prevent* [italics added] flashpoints. If I'd have had said nothing they might have just thought, well he hasn't seen it - we'll deal with it.

The other two from referee C were to prevent infringements from occurring, thus allowing the game to flow, for example:

There is a tackler on the wrong side and I was saying hands off to everyone who's got their hands on the ball. . . . I can't give a penalty if I'm not sure, so if I can work it so that that ball comes out then fine but as it is, it comes out, so I'm happy to play on.

Thus, when foul play and potential infringements are likely to occur, two of these referees communicate more firmly to players by preventative refereeing, enabling the game to flow yet remain in control.

#### *6.72 Increases in Hardness of Verbal and Non-Verbal Management*

Of the 42 incidents where the referees' management of the game became harder (medium or hard responses), 37 of these (88%) were deemed to be preventative, either to prevent a penalty, foul play or both. For example, referee B commented:

One of the reds comes in from the side. . . . Don't play advantage and my response is medium and you can tell that from my body language. I

wanted a quick penalty, because there's no way the ball would come back and if I allowed it to continue the feet would have come in from all over the place and it's called danger. . . . It could lead to injury, foul play, and potential red card.

In similar fashion referee C remarked:

I was obviously concerned about the blue tackler who got trapped. But again it's a quick whistle because I don't want boots on bodies. . . . So, I've gone for a quick whistle. . . . I don't want to have to go for a penalty for just getting too excited; hence me shouting out "leave him."

Later the same referee commented, "My response was medium because I'm having to continually talk to them to hold them back... to hold them back to prevent offside and also they've lost the ball and they desperately want to get it back." Although the remaining five cases were not tagged as preventative since there was no attempt to prevent a player from infringing the law, three were penalties (one of which was foul play meriting a yellow card) and the referee's condemnation of the player involved was a statement of intent to prevent further such actions. For example referee A stated:

It's big call it was 3 pts and it's the correct call – for the tackler not moving. . . . [Researcher: How hard was your response?] It was hard, and it was very specific. I'm pointing out to everyone that he's the reason that there's a penalty, and I'm also without actually saying he's gonna be yellow carded, I'm also saying to him, do not get involved again, that's two offences of the same nature in a game where we haven't had many penalties at all, and he's immediately replaced by the coach. . . . We've just had recently, a penalty for not releasing and this is another one for not releasing and it's a pattern that I don't want to see.

The other two were managed in a similar manner as the referee felt the temper of the game had risen, with referee C remarking:

I've got to be very clear as it is a period of pressure and the reds have just tried too hard which is why he's not released. It's very competitive – it's not dirty or nasty or anything like that – it is very competitive though.

So, once again, referees highlighted considerable use of preventative techniques in addition to the occasional need to be firm when serious infringements occurred, such as foul play; in this case, a unanimous finding.

Reviewing the position on the pitch, revealed that when the referees' management of the game became harder (medium or hard responses), 31 of these 42 incidents (74 %) were in the attacking half within kicking (i.e., penalty) range (see Table 6.3), for example referee A explained:

The defensive player was on the wrong side. . . . The anti's moved up. The defensive line has been moved back 20 meters through 2 to 3 phases, so there's space and potential opening up there. You're moving up the pitch as well, and the attack could turn into a try scoring opportunity. There's potentially more interest for the defenders to stop it developing, so my response goes to medium.

Similarly, remarking on another tackle near to the try scoring line, referee B said:

This guy came in from side. He didn't listen to me. . . . Deliberately killing the ball. It was a medium response because greens should've scored here and that's what I was looking at. . . . Very hair raising because it's in the red-zone. Red-zone's an area 10 metres from the try line where the defence will be quite happy to commit the 3 points for the penalty so that it stops the opposition getting 7 points [for a try].

Of the six medium/hard responses not in the attacking half for referee C, one was an incident of foul play and the other five were all either incidents where potential foul play might occur. In all cases the referee described his management as preventative.

Hence, although the pattern is not completely consistent for all three referees the precedent seems to be to manage harder when the play is more likely to lead to a scoring opportunity. Similarly, referee A and referee C's management was harder in the attacking 22m than in the rest of the field (referee A: hardness in 22m  $M = 2.62$ , non-22m hardness  $M = 2.07$ ; referee B: 22m hardness  $M = 1.29$ , non-22m hardness  $M = 1.48$ ; referee C: hardness in 22m  $M = 2.29$ , non-22m hardness  $M = 2.11$ ). These same two referees also managed harder (by mean level of response and frequency of medium & hard responses) in the second half than the first, though not significantly and this was not related to the scoreline at the time (all referees: 1<sup>st</sup> 15 minutes = 9 med/hard, 2<sup>nd</sup> 15 minutes = 14 med/hard, last 15 minutes = 19 med/hard). Although the "position on the pitch" trends as noted earlier were clearly related to the potential for scoring opportunities there were no consistent trends in the management of the game when the scoreline was close.

Finally, none of the referees made any reference to the weather or pitch conditions. Importantly, all three case studies were conducted on dry days with all the pitches in good condition; so unfortunately, this investigation offered no opportunity to consider this construct.

*Table 6.3 - Hardness of Referees' Responses (From 1-4) Comparing Kickable and Non-Kickable Tackles*

		Hardness of response						
		N of tackles	Mean	None %	Soft %	Med %	Hard %	
Referee A	98	62	2.32	31	47	21	2	Kickable
		36	1.83	36	61	3	0	Not kickable
Referee B	72	35	1.44	63	26	11	0	Kickable
		37	1.45	68	24	8	0	Not kickable
Referee C	78	50	2.24	14	60	12	14	Kickable
		28	1.96	32	43	21	4	Not kickable
Mean				33	46	16	5	
				47	43	10	1	

Mean hardness for all kickable tackles = 2.00

Mean hardness for all non-kickable tackles = 1.75

Percentage of medium & hard responses when kickable tackles (all referees) = 21.1%

Percentage of medium & hard responses when non-kickable tackles (all referees) = 10.9%

## 6.8 Phase Two Discussion

The second phase of this investigation revealed a fairly consistent picture of preventative refereeing. There is reasonable evidence that the factors identified in

phase one are indeed accounted for on the field of play, although it must be acknowledged that not all the referees exhibited all the factors. However, all three referees consistently managed harder as the play became more illegal and similarly they all became firmer as the potential for a scoring opportunity increased. Referees A and C presented more consistencies in their PMMs to referee B, which could be attributed to individual style or simply that his game did not require the same style of management. Finally, the richness of the responses from this elite group of referees showed a broad DMM base.

## 6.9 General Discussion

The purpose of this two-phase investigation was to establish the factors that affect referee DM in open team sports. The results revealed a consistently reported list of contextual factors, which were confirmed to be manifest in preventative refereeing; specifically, to prevent players from either infringing the law or committing foul play that may necessitate a more serious sanction. From a theoretical perspective, elite referees revealed a broad DMM base to interpret the context of the game, such as the likelihood of the opposition gaining a significant advantage from the situation and used their PMM skills to manage each situation based on this context; although there seemed to be some individual PMM styles. Referees were reluctant to award penalties when they are able to allow the game to flow by directing players away from infractions. These management skills seem to be crucial at the top level of such free flowing games.

### *6.91 Issues of Trustworthiness, Reliability and Validity*

Before drawing any conclusions it is important to acknowledge the limitations and consider how the nature of the investigation may have contributed to the findings. For example, there is evidence to suggest that think-aloud protocols encourage participants to describe what they think the researcher wants them to report (Nisbett & DeCamp Wilson, 1977). Nevertheless, I felt that the number of steps taken to ensure the accuracy of the data was sufficient enough to provide reliability, credibility, and trustworthiness. In this case the task was highly cognitive and the participants were regularly required to verbalise their decisions. In addition, I explored several data sources (two focus groups and three case studies) to increase the chances of valid findings (Fielding & Fielding, 1986). Also, participants in the focus groups developed the contextual factor themes rather than using a group of independent researchers who would typically not be present at the discussions. It was felt that this would help to ensure accurate categorising of the factors through ongoing interactions amongst the group. Subsequently, member checks were used to verify the accuracy of the findings and to give participants a further chance to amend their responses (see Patton, 2002).

In the second phase, two researchers independently examined the transcriptions to consensus validate the data. If any discrepancies were found between us, the video recordings of participant's responses were reviewed, capturing intonations and non-verbal communications, thus increasing the chances of accurately interpreting each

referee's meaning. Finally, I had already achieved a level of trust and rapport with the participants, increasing the chances of honest and reliable data (Fontana & Frey, 2000).

### *6.92 Implications*

Before presenting any implications to other research it is important to recognise the characteristics of rugby union investigated here. Rugby union is characterised by continuous, free flowing passages of play where the only stoppages occur when either the ball runs out of play or the referee penalises a player for an infringement in law. In this sense, it differs greatly from gridiron football where there are regular breaks after each down. Thus, an important but additional part of the referee's role in more "open" sports, such as rugby union, basketball, soccer, field and ice-hockey, seems to be to allow a game to flow and punish only those infringements that put their opponents at a disadvantage or when neglecting them may lead to dangerous flashpoints. For example, the international rules of basketball were amended specifically to address this issue, asking referees to understand the concept of "advantage/disadvantage," whereby the "officials should not seek to interrupt the flow of the game unnecessarily in order to penalise personal contact that is incidental and which does not give the player responsible an advantage nor place his opponent at a disadvantage" (FIBA, 2004). For this reason, referees use their understanding of the game's context (i.e., the tenor of the game, how much the players are listening to their requests, the potential consequence of awarding a penalty, and the potential for players taking the law into their own hands) to moderate their decisions. This



moderation or "weighting" appears to embrace potential effects on scoreline, tempered by the position on the pitch, potential flashpoints, often regulated by the temper of the game, all with an emphasis on steering players away from violations to allow the game to flow.

It is this management of the law and understanding of the spirit of the game that represents the more humanistic and artistic part of refereeing, contrasting with the more robotic application of the letter of the law. Unfortunately, many people involved in the game, especially the media, expect the referee to adhere to a more robotic model, demanding consistency (Metcalf, 2001; Weinberg & Richardson, 1990). Yet the results of this investigation clearly suggest that referees manage the game very differently depending on the context and when the potential consequences of sustaining an infringement may have a more significant impact on the game. They do not treat incidents in isolation but as a part of an unfolding series of events. It is this understanding of the game that elite referees should be striving for (Sabatini, 2002). Therefore, rather than applying a robotic application of the law or even "setting out your stall early on" by an excessively rigorous and tight application (cf. Weinberg & Richardson, 1990), refereeing at the top level is about "finding a set of solutions that work for you on the day," as propounded by referee A. Providing referees are fathomable to players and establish legitimacy in their decisions (Askins et al., 1981) the latitude in the laws remains one of the tools that referees have at their disposal to manage the game.

Given such clear yet potentially controversial findings, it appears that some education may be required to guide the media, players, coaches, fans and also NGBs in the way that referees apply the law. It seems that consistency may be an unrealistic demand thrust upon referees when the individual nuances of different games are accounted for, yielding fairness and legitimacy of decisions as a more sensible focus.

## 6.10 Conclusions

In conclusion, preventative refereeing is a crucial skill for elite referees demanding a broad DMM and PMM. Thus, assessments and training should be based on developing coherence in referees' mental models, rather than the "tick-box" approach to referee evaluations as propounded by the iRB. Allowing referees to establish a feel for the game based on player empathy, and understanding the consequences of potential decisions upon scoreline and player behaviour is crucial for elite referee development. Developing a broad repertoire of conflict management skills will allow referees to become more versatile, leading to free flowing games where the referee is not thrust into the centre of attention with regular and robotic whistle blowing. The longer the media and NGBs perpetuate the myth that refereeing is about making consistent decisions, referees will be cloned to become unfeeling and unthinking, robotic arbiters of the letter of the law. Unfortunately, until such changes occur, referees will be forced to either "referee to the boxes" in order to advance or

potentially risk hindering their career progression by refereeing an attractive, flowing game within the spirit of the sport.

## Chapter 7. Other Developments from RFU Support

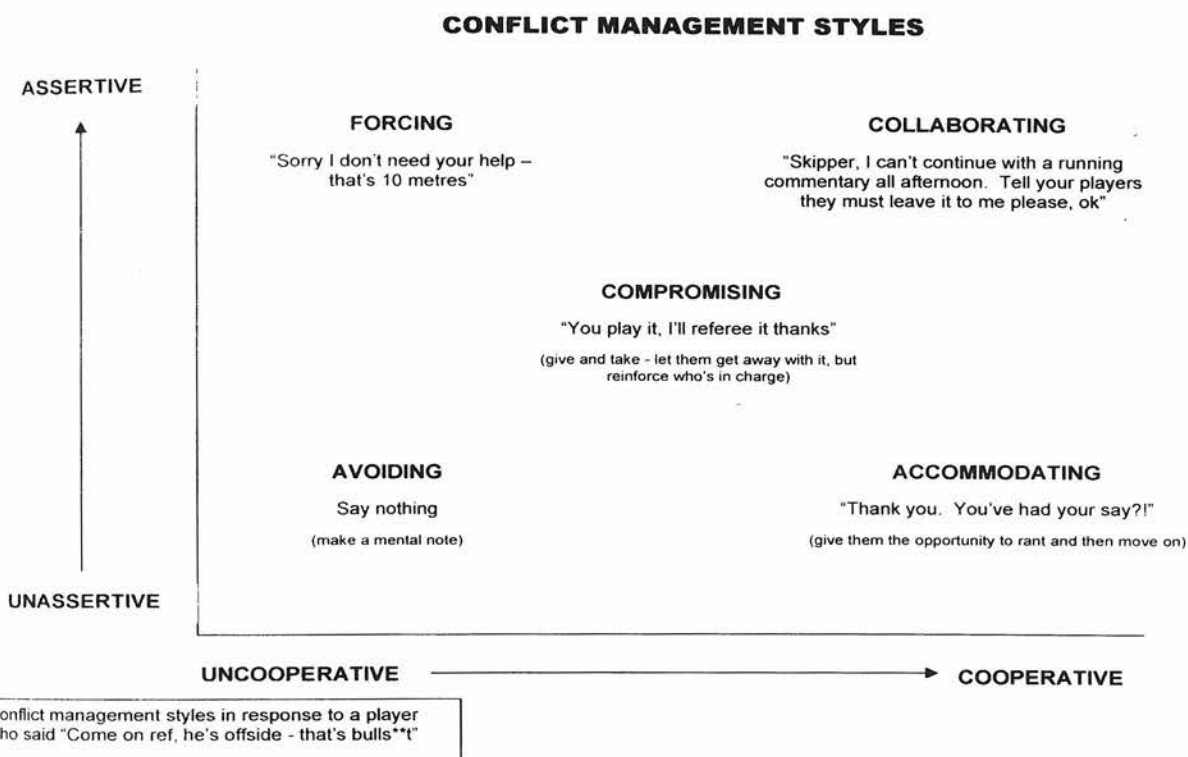
### 7.1 Personality & Management Skills Training

In attempting to understand, operationalise and intervene in each of the four base cornerstone areas many interventions took place. Just one example of such work, to assess and train personality and management skills, began with a review of literature on conflict management skills.

Utilising Thomas & Kilmann's (1974) conflict management style grid (see Figure 7.1), educational sessions were used to present video examples of different styles in refereeing, and 1-on-1 sessions (accompanied by video recordings of the referee's performance) were used to demonstrate individual referee's preferred styles. The model was also made rugby union referee specific, with examples provided for each style of management. The emphasis was on developing, first, an awareness of individual style and second, a more versatile range of management styles.

The model was referred to throughout group discussions to examine appropriate styles of management with different players and in a variety of situations. Referees were informed of the tendency to resort to type when the pressure increases and were thus encouraged to experiment with different styles during games that demanded a low mental workload. Feedback was triangulated through the referee, referee coach and sport psychologist to help ensure valid observations.

Figure 7.1 Conflict Management Grid



Adapted from Thomas & Kilmann (1974)

Thus the national panel of referees had a tool for examining one aspect of their behaviour to develop the personality and management skills cornerstone. Once again this assisted in developing a clear and consistent approach through the development of a common language.

## 7.2 Subsequent Law Application Interventions

Following the first law-application training package in 1999 and in response to the iRB's law change regarding "players entering the tackle from the side," another training package was conducted with new up-to-date tackles. Once again the package was tailor-made, focusing on this new area of law and also reinforcing the priority

system developed in the first training programme with new tackles (cf. Christina & Bjork, 1991). Similarly, in 2001 the iRB changed the emphasis on support players at the tackle, once again prompting another law-application training package. Regardless of these changes in both law and emphasis a coherence re-test in 2001 showed a 15% improvement in referee's accuracy since the start of the programme in 1998. This was due to the multiplicative impact through not only video based training but also discussions, presentations and the development of a coherent language.

The improvement in RFU referee performance was acknowledged both inside and outside the RFU referees. For example premier league coach Dean Richards remarked, "the referees have clearly improved, it's just the touch-judges now" (D. Richards, personal communication, January 19, 2002). Suggestions such as this, and in response to technological changes occurring in the professional game, attention was turned to the touch-judges.

### 7.3 Touch-Judge Coherence Training

Innovations in the southern hemisphere to give support to referees, led to the touch-judges having a communication link to the referee through a radio-microphone in their flags. This meant that the touch-judges were able to "feed-in" appropriate information to help the referee with his or her decisions. However, there had been no formal training on the use of this equipment, in particular on developing an understanding of what to feed-in and when. So, once again adopting a naturalistic approach, touch-judge perspective video clips were recorded to develop a coherence

training programme for touch-judging. Again, the aim was to develop role clarity (Carron, Spink & Prapavessi, 1997) by exposing an expert's declarative and procedural mental model, in this case for touch-judging. Developing clear

As reasoned in Chapter Five, I used a high-status (see McCullagh, 1986) touch-judge (who was an ex-international referee) and asked him to think-aloud during the recorded passages of play. When he felt it appropriate to feed information in to the referee he did so by pressing his button on the flag and talking into the microphone, just as he would during a game. The training package consisted of two videos. An assessment tape, which simply had the touch-judge perspective video clips and a training tape presenting the same clips with the expert's (think-aloud) voice-over describing what he was doing and why, with a "picture-in-picture" video of the expert appearing at the top corner of the screen to depict when he was feeding-in to the referee.

At the inaugural Elite Referee Unit Conference in 2002 all the touch-judges and referees were shown a section of the training tape to familiarise themselves with the think-aloud procedure and given the opportunity to watch a clip of video whilst thinking aloud and feeding in appropriate information to the referee, before viewing the expert's example. At this stage the expert's exemplars were not considered to be the unconditional answers to each clip, although they did provide valuable insights into the processes that take place, based on the verbalisation of declarative knowledge. All the referees as well as the touch-judges completed this "cross-training", i.e., attempting to fulfil the role of other team members, as this style of

training has enhanced team-member knowledge regarding the task, responsibilities and coordination strategies (Baker, Salas, Cannon-Bowers & Spector, 1992).

Once again this was a clear example of how a coherent language was developed amongst the RFU officials. Furthermore, both qualitative and quantitative feedback (see Poczwadowski, Sherman & Henschen, 1998) from the 23 delegates was very positive. They rated the *value* of this section of the conference to be  $M = 9.5$  (on a Likert scale from 1-10), and provided comments such as:

“Extremely thought provoking and useful”

“Cleared up lots of issues that were not clear”

“Excellent, very tiring”

Other changes in the game prompted more touch-judge training.

## 7.4 Touch-Judge Communication Training

Similarly, in reaction to the increasing number of games on broadcast television and the microphone links to the team of officials, the RFU management team identified the need to develop a protocol for touch-judges reporting foul play to the match referees. Given that foul play incidents invariably created a break in play, turning attention towards the officials, there was a lot of pressure on the touch-judges to report very clearly, accurately and with confidence. To assist them they were trained to remember the acronyms CAC and RAC for their verbal and non-verbal responses. Verbally, their goal was to be:



- Clear,
- Accurate, and
- Concise

Non-verbally, the aim was to project a manner that was:

- Relaxed,
- Assured, and
- Confident

The emphasis was on refraining from using emotive or colloquial phrases such as “the geezer stamped right on his head, he’s got to go,” to less litigious language, such as “number five red - unnecessary use of the boot. I recommend a red card.”

In attempts to simulate the pressure experienced in the naturalistic environment (cf. Salas et al., 1997) a group of touch-judges were asked to stand in front of their peers, watch an incident of foul play from touch-judge perspective video (cf. Omodei et al., 1998) and were given approximately 20 seconds (as they would experience in a real game) to prepare their dialogue for the referee. A touch-judge was selected at random and asked to deliver his response to a camcorder in front of a room full of his peers. The camera added to the pressure experienced and provided a good medium through which to feedback. As with the other training programmes, model answers were pre-recorded to generate further discussions and to increase coherence. Once again the delegate feedback on this session was extremely positive ( $M = 9$ , out of 10) and led to comments on the touch-judge training conference such as:

“Without a doubt the most productive of all the conferences I have attended.”

“Excellent from both a personal and development perspective. Made to be both hard work, informative and enjoyable.”

“As a critic of conferences in the past this was an excellent conference, with the Edinburgh University facilitators doing a first-class job.”

The development of both of these touch-judge training programmes reflected the need to maintain an R&D culture, flexible and adaptive to changing demands and thus provide sound, evidence-based practice.

## 7.5 Organisational Issues

During the period of support, the RFU underwent considerable change to keep pace with the professional game. In 1999 the RFU employed three full-time professional referees (the first full-time referees in the UK), eventually increasing that number to five in 2001. Accordingly, sport science support was tapered to meet the increasing demands of the full-time referees, with regular 1-to-1 meetings and team training sessions at the referee centre of excellence.

Since the beginning of the research program all the administrative and financial aspects of refereeing had been conducted through the centre of excellence under the auspices of the referee department. In July 2001, the Elite Referee Unit was formed, with sole responsibility for officiating at the premier league level. At this time the top 15 referees, 10 touch-judges, five assessor coaches, and three management staff moved from the referee department to the newly formed elite group. Reflecting how

sport psychology support seems to be changing (Jones, 2002), I had considerable input to the organisational development of the RFU.

To aid the transition to the new elite referee unit, the sport psychology team assisted with organisational psychology input, providing “change management” advice and support. We assisted in creating an environment with open systems for selection (Yukelson, 1997) and developed a formal process of promotion and relegation to and from the elite group. In addition, the sport science team became a key feature of the new elite group, whilst still maintaining links with the referee department to help ensure that an impact was made across all levels.

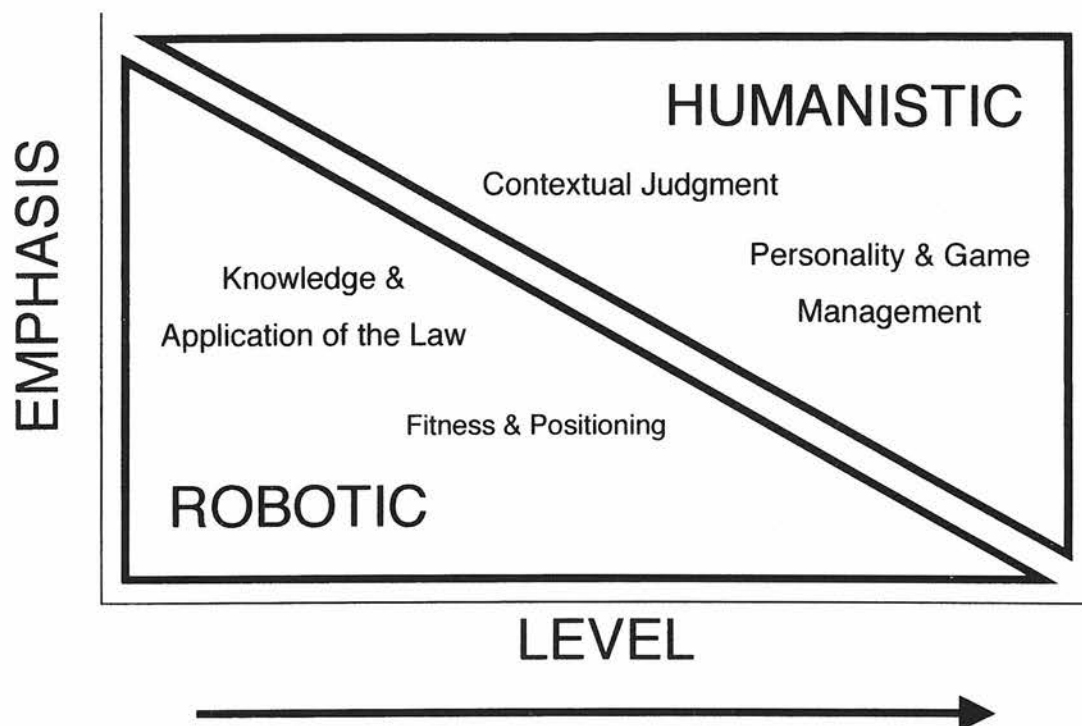
I conducted organisational psychology support throughout the project through regular attendance at meetings and 1-on-1 support for the management group. This typically included long and short-term goal setting with regular performance reviews designed to enable the RFU staff to become more efficient and stress free. Regular contact with this group also led to developments in the understanding of the characteristics highlighted in the performance model.

## 7.6 Ongoing Development of the Performance Model

Following observations from the RFU management staff about the robotic and humanistic nature of elements in the performance model (identified in Chapter Two), I was interested in developing these ideas. For example, subsequent discussions with the RFU managers suggested that the emphasis for support in the humanistic areas was not only more difficult but as a consequence typically developed much later in

referees' careers. Building on these ideas the following "Training Emphasis Model" emerged (Figure 7.2).

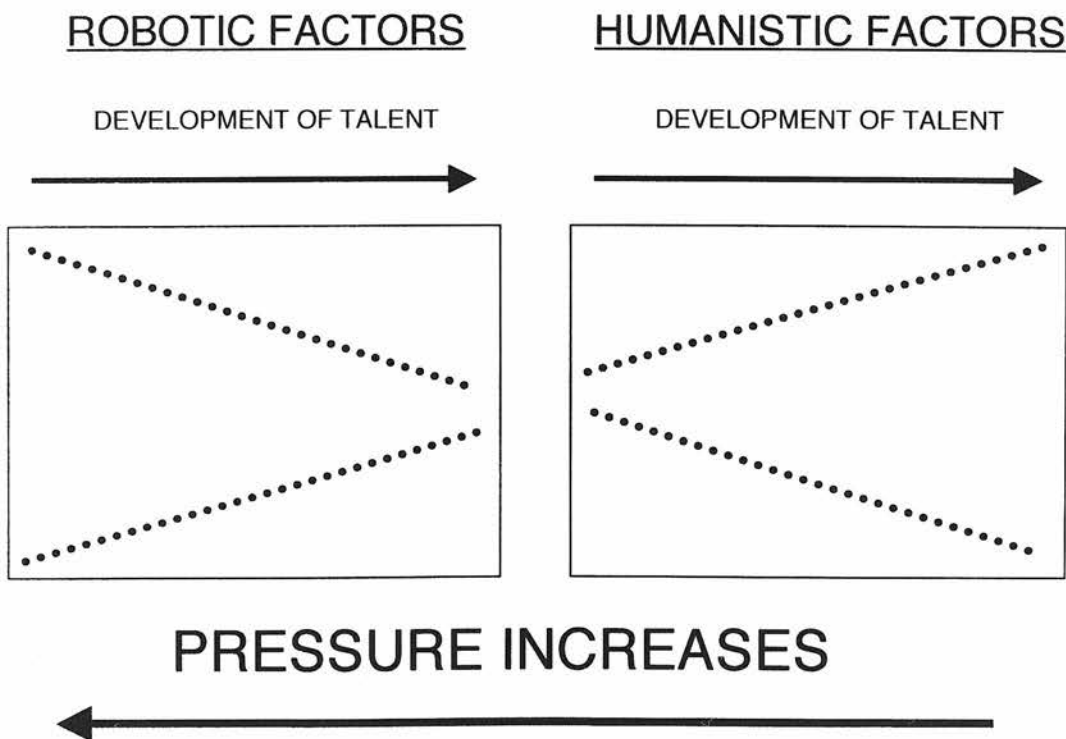
*Figure 7.2 Training Emphasis Model on the Robotic and Humanistic Skills of Refereeing*



This model suggests that as referees move to higher performance levels the training emphasis shifts from robotic towards the more humanistic factors. It should be accepted that referees at the highest levels have a good knowledge and understanding of the application of law, yet (as suggested in Chapter Five) their contextual judgment skills may still need development as they become more important. The reverse is true for the less experienced referees new to refereeing, who should primarily focus on developing their levels of fitness and positioning, and how they

apply the law. Correspondingly, “ProZone” a notational analysis package, has recently shown that the soccer referee considered by many to be the world’s best, Pierre Luigi Collina (Henderson, 2002), uses contextual judgment and knowledge of the game to reduce the need for the highest levels of physical fitness by not entering the penalty box at either end of the field (Hackett, 2005). Hence, while each corner of the performance model is crucial for successful refereeing performance (see Anshel, 1995) not only may the requirement shift as referees’ progress, but referees also seem able to compensate for weaknesses in some areas by emphasising others. In similar vein to the “Training Emphasis Model,” research into the pressure experienced by referees led to another model emerging that was used as an educational tool for referees (see Figure 7.3).

*Figure 7.3 Talent Development and Responses to Pressure in Refereeing Skills*



This model highlighted the call for reducing the degrees of freedom in the robotic factors and increasing the degrees of freedom in the humanistic factors. For example, all referees should be aspiring to apply the law coherently so that they are consistent both within themselves and between each other, in effect decreasing the spread of their decisions. Additionally, referees should be aiming to broaden their repertoire of management skills so that they develop a diverse range of styles and can subsequently adapt to refereeing an array of scenarios (Topp, 1999).

Crucially, the model shows that when referees experience stress, their performance is likely to do the opposite; their law application is likely to become more widespread and they are likely to resort to type in their style of management. This model was a valuable reference during individual referee support sessions.

## 7.7 Elite Referee Support

Much of the individual support work with the elite referee group centred on developing the psychological characteristics of excellence that overarch the performance model. In support of the interactive nature of the key cornerstones in the model, this work often spanned different cornerstones. For example, periodised performance plans were conducted to ensure referees reached their physical and mental peak at key periods during the year and physically challenging fitness training regimes were undertaken at key times to develop mental toughness through the physical toughness to mental toughness link (Loehr, 1995). This required an

integrated approach, between the physiologist and psychologist. Finally, adopting a critical and reflective approach to support work assisted all the psychology team towards ensuring that they were having maximum impact (Anderson, Knowles & Gilbourne, 2004).

## 8. Summary, Conclusions and Future Recommendations

### 8.1 Introduction

This thesis described an R&D based support programme with the RFU referees. The program lasted 5 years, which was testament in itself to the efficacy of the support. This chapter summarises each part of the thesis, providing the rationale for each, how I approached each investigation, highlighting the key findings and implications, and outlines future directions for research.

### 8.2 Summary of Thesis

Chapter One introduced the RFU referee scientific support program and highlighted the need for a better understanding of referee performance due to unclear guidelines from NGBs and a dearth of empirical research. It proposed a research programme to first, understand what constitutes elite refereeing performance; second, explore methods to operationalise those key areas; and third produce empirically driven interventions to develop each area.

Consequently, Chapter Two examined RFU data, scientific research and performance profiles of top-ranked referees to develop a refereeing performance model. The Cornerstones Performance Model of Refereeing emerged, featuring four base cornerstones: comprising knowledge and application of law; contextual judgment; personality and game management; and fitness and positioning; all



overarched by the psychological characteristics of excellence (see McCaffrey & Orlick, 1989).

Given the significance of the DM elements of refereeing performance, in Chapter Three different scientific approaches were reviewed, including literature on motor control, judgments, and classical and naturalistic decision making. The naturalistic approach was adopted as it centres on the examination of experts operating in highly stressful environments, congruent to the arena of open sports officiating.

Recognising the importance of law application (Anshel, 1995) and in response to RFU demands, Chapter Four described a video assessment tool developed to test the accuracy and coherence of the RFU referees and their support personnel. In accordance with NDM theory, the tool was designed to be as ecologically valid as possible, taking typical scenarios from real game situations. The test revealed only 50% accuracy from referee groups and even poorer results from their support groups, suggesting the need for a structured training programme.

Accordingly, Chapter Five reported the development of a law-application training tool designed to improve the accuracy and coherence of the RFU referees by exposing an expert's SMM. Lower ranked referees on the national panel improved their performance by 17% from pre to posttest. This provided the RFU referees with a scientifically validated training tool to conduct cognitively based, deliberate practice off the field of play. Despite such positive results, even the top referees were

still only achieving approximately 50% accuracy, so it was suggested that contextual judgment might supersede the application of law.

Consequently, Chapter Six used two focus groups to explore other factors that might affect referees' application of law. Subsequently, a think-aloud protocol was conducted with three international referees, to investigate the affect of context on refereeing. These elite referees were found to use a preventative refereeing style, particularly when the consequences on scoreline and negative player behaviour increased. Such findings highlighted the need for more appreciation of this crucial area and the development of interventions to train both declarative and procedural skills.

Highlighting the need for a complete and coherent work programme Chapter Seven describes other aspects of support work that were carried out. For example, in response to changing demands two interventions to improve touch-judges core skills were described. An expert's declarative and procedural mental model for touch judging was exposed using high quality video scenarios. Very positive feedback suggested that the training was a success, reinforcing the need for a flexible approach to scientific support.

### 8.3 The RFU Legacy

Developing close working relationships throughout the RFU was a key to the success of this project. Clearly this was important for individual referee support, but ongoing attendance and presentations at meetings allowed me to keep abreast of arising issues

in the game and respond accordingly. This created a flexible model of practice, capable of making an impact throughout the organisation, not only at the top-level but also at the grass roots through training products that were developed. This included:

- Three law application coherence training packages.
- A touch-judge coherence training package.
- A touch-judge communication training package.
- An RFU referee agility training video.

Adopting a scientific approach has made an impact through quantifiable changes in performance as well as many improvements that cannot be measured. For example, the project has also affected the way the RFU referees think and talk about performance. The project has given them a language, where terms such as “coherence” and “shared mental models” are commonplace and seen as an important part of developing their performance. This was supported by more formal appraisals too. Yearly evaluations with the Consultant Evaluation Form (Partington & Orlick, 1987) routinely indicated generally positive responses and the overall perceptions were that I had a beneficial effect on referee performance.

Interestingly, two development referees who were identified as having high potential by criterion in the Cornerstones Performance Model of Refereeing are now full-time RFU referees operating at premiership and junior international level. Furthermore, the coaching forms based on this model are still used right across the RFU and perhaps most significantly, three of the referees who have received support

throughout this project have now progressed to the international A-list (world top 16), one of whom was short listed to referee the world cup final in 2003 (Iles, 2004).

## 8.4 Contributions to the Body of Knowledge

In addition to applied contributions my research has formed a crucial part of the emerging body of knowledge, as the emphasis has changed to performance-based research (e.g., Oudejans et al., 2002; Plessner & Betsch, 2001). My contribution is evidenced in debates with other researchers, for example a paper entitled “The Art of Reason versus the Exactness of Science in Elite Refereeing: Comments on Plessner & Betsch (2001)” was published in the *Journal of Sport and Exercise Psychology* (see Appendix Nine for the full reference).

This paper commended Plessner & Betsch’s (2001) study on referee DM but highlighted two methodological weaknesses and a crucial error in the attribution of their findings. Crucially, Plessner & Betsch failed to simulate the match-day referee’s perspective, forfeiting ecological validity (see Chamberlain & Coelho, 1993) and most importantly in failing to examine the reasons underpinning decisions (Abraham & Collins, 1998) made a series of erroneous claims.

As a consequence of this research paper, there has been a growing interest from governing bodies to support such research into crucial aspects of referee performance. The “New Zealand Sport and Recreation Agency,” funded by the New Zealand government, has recently commissioned a university project to assess the performance of their national league soccer referees. Specifically, this project will

look at the effect of movement patterns and physical workload upon the quality of their decisions. Similar projects look likely to develop in Germany, with funding from the German Football Federation (H. Plessner, personal communication, October 9, 2005).

In addition to these collaborations with other researchers, the work conducted in Chapter's Two, Four and Five have been accepted for publication and the performance model has received recognition from the (American) National Association of Sports Officials (see Appendix Nine for the full references). These have developed theoretical understanding by building upon the NDM paradigm. For example, Chapter Five highlighted the idea of developing coherence on three levels,

1. decision,
2. reason, and
3. explanation of that reasoning

to ensure that referees' SMMs are completely isomorphous.

In addition to a number of technical reports, workshops and presentations for the RFU (see Appendix Nine), new theoretical understanding has also been disseminated through contributions at academic conferences. For example, I presented at the British Psychological Society refereeing symposium at their national conference in 2003 and also the British Association of Sport and Exercise Sciences conference where I was the first-place prize-winner (see Appendix Nine). Collectively, it is hoped that such contributions will encourage similar referee performance research.

Finally, as a corollary of this research project with the RFU referees, the interventions have received interest from the English and Scottish Football Associations, the English Rugby Football League, the Scottish Basketball Association and Major League Soccer, resulting in training programmes across different sports.

## 8.5 Future Directions

Future assessments and interventions should continue this theme to establish appropriate levels of accuracy and coherence in different sports. Further research is also needed to develop even more ecologically valid approaches, for example, simulating the physical as well as cognitive demands experienced on the field.

The only published empirical research that has attempted to assess the physical and perceptual-cognitive demands of referees has been conducted in soccer (Helsen & Bultynck, 2004). However, by only assessing *observable* decisions, rather than employing a more cognitively orientated investigation, perhaps through the use of a think-aloud protocol, their findings may be somewhat erroneous. Their investigation claimed that soccer referees in the final round of the Euro 2000 Championship made only 137 observable decisions per game. Yet, measuring only *observable* decisions may not, by its very nature, be representative of *cognitive* decisions. Moreover, given the richness of the qualitative data in Chapter Six of this thesis, it seems that referees are constantly making decisions, often not directly on the play presented to them, but

on planning future scenarios. Thus, future research should consider mixed method designs to investigate the cognitive performance of sports officials.

In similar fashion, since this thesis has identified the importance of recognising the interactions between the corners of the Performance Model of Refereeing, an interdisciplinary approach to such research would be most beneficial.

Methodologically, the success of the naturalistic approach that has been adopted throughout this referee support program shows the value in exploring other fields of research for appropriate theoretical perspectives. As such, further explorations into technologies such as head-mounted video cameras that have been used extensively in NDM research (e.g., Omodei & McLennan, 1994) should be considered for investigating sports officials as they can produce high fidelity and real-time recordings of the naturally experienced environment (Omodei, Wearing & McLennan, 1997). Thus, as propounded by Omodei et al. (1997), this level of fidelity can increase the quality of stimulated-recall and think aloud protocols as explored in Chapter Six.

Furthermore, linking the findings from the think-aloud protocol in Chapter Six to the work carried out in Chapter Seven, referees' advanced declarative and procedural skills could be measured and trained with the use of the conflict management grid to plot approaches to conflict resolution when the context of the game changes, thus providing a quantifiable solution to measuring such humanistic skills; an approach

that NGBs would welcome, given countless attempts to develop quantitative tools to measure refereeing performance (e.g., Griffiths, et al., 1999).

Furthermore, given the severe financial implications of poor refereeing performance (Craven, 1998), a big challenge for refereeing NGBs is to find appropriate ways of introducing young, developing referees to officiating at the very top level (K. Hackett, General Manager of the FA Professional Game Match Officials' Board, personal communication, February 10, 2005). As such, explorations into a type of flight simulator for refereeing, capable of simulating the real world environment of refereeing, should reflect upon the plethora of work in NDM literature that investigates decision making in aviation, (e.g. Orasanu & Fischer, 1993), and also physiological assessments of referee performance (e.g., Catterall, Reilly, Atkinson & Coldwells, 1993) perhaps through the use of treadmills and widescreen television projections.

Finally, Chapter Seven revealed some potentially controversial findings that seem to oppose the demand for *consistent* refereeing, highlighting the need to develop training in this area for elite referees in rugby union. Correspondingly, more research is required to see if the same findings would be apparent in other open team sports. Anecdotal evidence suggests that elite soccer referees do indeed use similar preventative game management skills to referee games. Reflecting on a recent Manchester United versus Arsenal FA premiership game, referee Graham Poll stated that by the letter of the law, two players should have been sanctioned with yellow cards in the tunnel before the start of a game, but "I chose to manage the situation



because I thought it was best for the game” (Graham Poll, FA premiership referee, personal communication, Feb. 10, 2005). Thus, further scientific examination is required.

## References

- Abernethy, B. (1996). Training the visual-perceptual skills of athletes. Insights from the study of motor expertise. *American Journal of Sports Medicine*, 24(suppl. 6), S89-S92.
- Abernethy, B., & Russell, D. G. (1987). The relationship between expertise and visual search strategy in a racquet sport. *Human Movement Science*, 6, 283-319.
- Abernethy, B., Thomas, K. T., & Thomas, J. T. (1993). Strategies for improving understanding of motor expertise (or mistakes we have made and things we have learned). In J. L. Starkes, & F. Allard (Eds.), *Cognitive issues in motor expertise* (pp 317-356). Amsterdam, Netherlands: Elsevier.
- Abernethy, B., Wood, J. M., & Parks, S. (1999). Can the anticipatory skills of experts be learned by novices? *Research Quarterly for Exercise and Sport*, 70(3), 313-318.
- Abraham, A., & Collins, D. (1998). Examining and extending research in coach development. *Quest*, 50, 59-79.
- Ackford, P. (2002, May 25). Leicester reign again as Back robs Munster of glory. *The Daily Telegraph*.
- Ackford, P. (2003, March 16). Ring of confidence from the whistle blowers. *The Sunday Telegraph*, p. 11.
- Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction*, 15(2), 40-47.
- Altman, D. G. (1991). *Practical statistics for medical research*. Boca Raton, FL: Chapman and Hall.

- American Psychological Association. (2003). *Publication manual of the American Psychological Association* (5<sup>th</sup> ed.). Washington, DC: American Psychological Association.
- Anderson, A. G., Knowles, Z., & Gilbourne, D. (2004). Reflective practice for sport psychologists: Concepts, models, practical implications, and thoughts on dissemination. *The Sport Psychologist*, 18, 188-203.
- Anderson, J.R. (1995). *Cognitive psychology and its implications* (4th ed.). W. H. Freeman and Company: New York.
- Annett, J. (1997). Analysing team skills. In R Flin, E Salas, M Strub, & L Martin (Eds.), *DM under stress: Emerging themes and applications* (pp. 315-325). Aldershot: Ashgate.
- Anshel, M. H. (1995). Development of a rating scale for determining competence in basketball referees: Implications for sport psychology. *The Sport Psychologist*, 9, 4-28.
- Anshel, M. H., & Webb, P. (1991). Defining competence for effective refereeing. *Sports Coach*, 14(3), 32-37.
- Anshel, M. H., & Weinberg, R. S. (1999). Re-examining coping among basketball referees following stressful events: implications for coping interventions. *Journal of Sport Behavior*, 22(2), 141-161.
- Askins, L., Carter, T., & Wood, M. (1981). Rule enforcement in a public setting: the case for basketball officiating. *Qualitative Sociology*, 4(2), 87-101.
- Askins, R. (1987). Common myths about officiating *Referee*, 26(10), 44-47.

Baker, C. V., Salas, E., Cannon-Bowers, J. A., & Spector, P. (1992, April). *The effects of inter-positional uncertainty and workload on teamwork and task performance*. Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, Montreal, Canada.

Balke, W. M., Hammond, K. R., & Meyer, G. D. (1973). An alternate approach to labor-management negotiations. *Administrative Science Quarterly*, 18, 311-327.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, N.J: Prentice-Hall.

Barwick, B. (Executive Producer) (2003, November 22) *Grandstand: 2003 rugby union world cup*. London: British Broadcasting Corporation.

Beach, L. R., & Lipshitz, R. (1993). Why classical decision theory is an inappropriate standard for evaluating and aiding most human decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok, (Eds.), *Decision making in action: models and methods* (pp. 21-35). New Jersey: Ablex Publishing Corporation.

Bloom, B. S. (1953). Thought-processes in lectures and discussions. *Journal of General Education*, 7(3), 160-169.

Bloom, B. S. (1985). *Developing Talent in Young People*, New York, Ballantyne.

Brannick, M. T., & Prince, C. (1997). An overview of team performance measurement. In M. T. Brannick, Salas, E., & Prince, C. (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 3-16). New Jersey: Lawrence Erlbaum.

Brehmer, B. (1972). Policy conflict as a function of policy similarity and policy complexity. *Scandinavian Journal of Psychology*, 13, 208-221.

- Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, 81, 211-241.
- Bull, S. J. (1989). The role of the sport psychology consultant: A case study of ultra-distance running. *The Sport Psychologist*, 3, 254-264.
- Bunting, N. J. (1998). Rugby Football Union Refereeing. *Welcome to the national conference, Bromsgrove, July 1999*.
- Bunting, N. J. (1999). *Allied Dunbar premiership guidance for referees, players and coaches on the application of law* (RFU Tech. Rep. from the conference on the game). Castlecroft, Wolverhampton.
- Burke, K. L., Joyner, A. B., & Pim, A. (2000). An exploratory investigation of the perceptions of anxiety among basketball officials before, during, and after the contest. *Journal of Sport Behavior*, 23(1), 11-19.
- Butler, R. J., & Hardy, L. (1992). The performance profile: Theory and application. *The Sport Psychologist*, 6, 253-264.
- Calderwood, R., Crandall, B. W., & Klein, G. A. (1987). *Protocol analysis of expert/novice command decision making during simulated fire ground incidents*. Yellow Springs, OH: Klein Associates. Prepared under contract for the U.S Army Research Institute.
- Calderwood, R., Klein, G. A., & Crandall, B. W. (1988). Time pressure, skill and move quality in chess. *American Journal of Psychology*, 101, 481-493.
- Campsall, B. (2002). Refereeing the Tackle. In High, C. J. (Chair), Rugby Football Union Performance Department: Inaugural Elite Referee Unit Conference, *Huddersfield, August 2002*.

Cannon-Bowers, J. A., & Bell, H. H. (1997). Training decision-makers for complex environments: Implications of the naturalistic decision-making perspective. In C. E. Zsombok, & G. Klein (Eds.), *Naturalistic decision making* (pp. 99-110). Mahwah, NJ: Lawrence Erlbaum.

Cannon-Bowers, J. A. Salas, E., & Converse, S. A. (1990). Cognitive psychology and team training: Shared mental models in complex systems. *Human Factors Society Bulletin*, 33(12), 1-4.

Cannon-Bowers, J. A. Salas, E., & Pruitt (1996). Establishing the boundaries of a paradigm for decision making research. *Human Factors*, 38(2), 193-205.

Carron, A., Spink, K. S., & Prapavessi, H. (1997). Team building and cohesiveness in the sport and exercise setting: Use of indirect interventions. *Journal of Applied Sport Psychology*, 9, 61-72.

Catterall, C., Reilly, T., Atkinson, G., & Coldwells, A. (1993) Analysis of the work rates and heart rates of association football referees. *British Journal of Sports Medicine*, 27(3), 193-196.

Christina, R. W., & Bjork, R. A. (1991). Optimizing long-term retention and transfer. In Druckman, D., & R.A. Bjork (Eds.). *In the mind's eye: Enhancing human performance* (pp. 23-56). Washington, D.C: National Academy Press.

Clark-Carter, D. (2001). *Doing quantitative psychological research: From design to report*. East Sussex, UK: Psychology Press.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences* (2<sup>nd</sup> ed.). Hillsdale, New Jersey: Lawrence Erlbaum.

- Cohen, M. S. (1993). The naturalistic basis of decision biases. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 51-94). New Jersey: Ablex Publishing Corporation.
- Cohen, M. S., Freeman, J. T., & Thompson, B. B. (1997). Training the naturalistic decision maker. In C. E. Zsombok, & G. A. Klein, (Eds.), *Naturalistic decision making* (pp. 257-268). New Jersey: Lawrence Erlbaum.
- Converse, S. A., & Kahler, S. E. (1992). *Knowledge acquisition and the measurement of shared mental models*. Unpublished manuscript, Naval Training Systems Center, Orlando, FL.
- Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring team knowledge. *Human Factors*, 42(1), 151-173.
- Craven, B. J. (1999). A psychophysical study of leg-before-wicket judgments in cricket. *British Journal of Psychology*, 89, 555-578.
- de Groot, A. (1965). *Thought and choice in chess*. The Hague: Mouton.
- Delbecq A., & VandeVen, A. (1971). A group process model for problem identification and program planning. *Journal of Applied Behavioral Science*, 7, 466-491.
- Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Operations Research Center, University of California, Berkeley, ORC-80-2.
- Druckman, D., & Bjork, R. A. (1991). Modeling expertise. In D. Druckman & R. A. Bjork, (Eds.), *In the mind's eye: Enhancing human performance* (pp. 57-79). Washington, DC: National Academy Press.

Eaves, S., & Hughes, M. (2003). Patterns of play of international rugby union teams before and after the introduction of professional status. *International Journal of Performance Analysis in Sport*, 3(2), 103-111.

Endsley, M. R. (1997). The role of situation assessment in naturalistic decision making. In C. E. Zsombok, & G. A. Klein, (Eds.), *Naturalistic decision making* (pp. 269-283). New Jersey: Lawrence Erlbaum.

Ericsson, K.A. (2003). Development of elite performance and deliberate practice. In J. L. Starkes, & K. A. Ericsson, (Eds.), *Expert performance in sports: Advances in research on expertise* (pp. 49-83). Champaign, IL: Human Kinetics.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence for maximal adaptations on task constraints. *Annual Review of Psychology*, 47, 273-305.

Eylon, B. S., & Reif, F. (1984). Effects of knowledge organization on task performance. *Cognition and Instruction*, 1, 5-44.

Farrow, D., & Abernethy, B. (2002). Can anticipatory skills be learned through implicit video-based perceptual training? *Journal of Sports Sciences*, 20, 471-485.

FIBA – Fédération Internationale de Basketball (2004). *Official basketball rules 2004: Referees' manual for two-person officiating*. Paris, France: FIBA.

Fielding, N. G., & Fielding, J. L. (1986). *Linking data*. Beverly Hills, CA: Sage.

Fontana, A., & Frey, J. H. (2000). The interview: From structured questions to negotiated text (pp. 645-672). In N. Denzin, & V. S. Lincoln, (Eds.), *Handbook of qualitative research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications.



- Ford, G. G., Gallagher, S. H., Lacy, B. A., Bridwell, A. M., & Goodwin, F. (1999). Repositioning the home plate umpire to provide enhanced perceptual cues and more accurate ball-strike judgments. *Journal of Sport Behavior*, 22(1), 28-44.
- Franks, I. M., Elliott, M., & Johnson, R. (1985). *The effects of experience on the detection and location of performance differences in a gymnastic technique*. Paper presented at the meeting of the Canadian society for psychomotor learning and sport psychology, Montreal, Canada.
- Garcia, A. (n.d.). *The professional boxing referee: Evaluating the professional boxing referee - part one*. Retrieved December 11, 2003, from [http://www.secondsout.com/ringside/goodman\\_45692.asp](http://www.secondsout.com/ringside/goodman_45692.asp)
- Gentner, D., & Stevens, A.L. (1983). *Mental models*. New Jersey: Lawrence Erlbaum.
- Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39, 93-104.
- Goldsmith, P. A., & Williams, J. M. (1992). Perceived stressors for football and volleyball officials from three rating levels. *Journal of Sport Behavior*, 15, 106-118.
- Griffiths, S., Dickinson, M., & Renton, P. (1999). Competence based referee development system. *RFU Referees Report July 1999*. Castlecroft, Wolverhampton.
- Grunska, J. (1999). *Successful sports officiating*. Champaign, IL: Human Kinetics.
- Hackett, K. (2005) [Using "ProZone" for analysing Referee work rates and movement patterns] Unpublished raw data.
- Hardy, L., & Parfitt, G. (1994). The development of a model for the provision of psychological support to a national squad. *The Sport Psychologist*, 8, 126-142.

- Hayes, (2000). *Doing psychological research*. Suffolk, England. Open University Press.
- Heath, S. B. (1991). "It's about winning!" The language of knowledge in baseball. In L. Resnick, J. Levine, & S. Behrend (Eds.), *Shared cognition: Thinking as social practice*. Washington, DC: American Psychological Association.
- Helsen, W. F., Hodges, N. J., Van Winckel, J., & Starkes, J. L. (2000). The roles of talent, physical precocity and practice in the development of soccer expertise. *Journal of Sport Sciences*, 18, 727-736.
- Helsen, W., & Pauwels, J. M. (1993). The relationship between expertise and visual information processing in sport. In J. J. Starkes, & F. Allard, (Eds.), *Cognitive issues in motor expertise* (pp. 109-134). Amsterdam, Netherlands: Elsevier.
- Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport and Exercise Psychology*, 20, 12-34.
- Henderson, J. (2002, May 19). Whistle while you work. *The Observer*.
- Holyoak, K. J. (1984). Mental models in problem solving. In J. R. Anderson, & S. M. Kosslyn. (Eds.), *Tutorials in learning and memory* (pp. 103-218). New York: W. H. Freeman and Company.
- Hutchins, S. G. (1997). Decision-making errors demonstrated by experienced naval officers in a littoral environment. In C. E. Zsombok, & G. A. Klein, (Eds.), *Naturalistic decision making* (pp. 207-215). Mahwah, NJ: Lawrence Erlbaum Associates.
- Iles, R. (2004). *The life of rugby referee Chris White*. Retrieved October 10, 2005, from [http://www.bbc.co.uk/gloucestershire/focus/2004/03/chris\\_white.shtml](http://www.bbc.co.uk/gloucestershire/focus/2004/03/chris_white.shtml).

iRB – international Rugby Board. (1995). *The laws in plain English*. Bristol, England: IRFB Services Ltd.

iRB - international Rugby Board. (2003). *The laws of the game of rugby union. (2003 ed.)*. Retrieved December 11, 2003, from [http://www.irb.com/laws\\_regs/laws/index.cfm](http://www.irb.com/laws_regs/laws/index.cfm)

Isenberg, D. J. (1985). Some hows and whats of managerial thinking: Implications for future army leaders. In J. G. Hunt, & J. Blair (Eds.), *Military leadership in the future battlefield*. New York: Pergammon Press.

Johnson-Laird, P. N. (Ed.) (1983). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

Johnston, J. H., Poirier, J., & Smith-Jentsch, K. A. (1998). Decision making under stress: Creating a research methodology. In J. A. Cannon-Bowers, & E. Salas, (Eds.), *Decision making under stress: Implications for individual and team training* (pp. 39-59). Washington, DC: American Psychological Association.

Jokela, M., & Hanin, Y. L. (1999). Does the individual zones of optimal functioning model discriminate between successful and less successful athletes? A meta-analysis. *Journal of Sport Sciences*, 17(11), 873-887.

Jones, G. (2002). Performance excellence: A personal perspective on the link between sport and business. *Journal of Applied Sport Psychology*, 14, 268-281.

Jones, M. V., Paull, G. C., & Erskine, J. (2002). The impact of a team's aggressive reputation on the decisions of association football referees. *Journal of Sports Sciences*, 20, 991-1000.

Jones, S. (2005, May 15). Woodward sees red over lions tour officials. *The Sunday Times*, (2-17).

Kaissidis-Rodafinos, A., Anshel, M., & Porter, A. (1997). Personal and situational factors that predict coping strategies for acute stress among basketball referees. *Journal of Sport Sciences, 15*, 427-436.

Kaissidis-Rodafinos, A., Anshel, M. H., & Sideridis, G. (1998). Sources, intensity, and responses to stress in Greek and Australian basketball referees. *International Journal of Sport Psychology, 29*, 303-323.

Kerr, J. H. (1999). The role of aggression and violence in sport: A rejoinder to the ISP position stand. *The Sport Psychologist, 13*, 83-88.

Klein, G. A. (1989). Recognition-primed Decisions. In W. B. Rouse (Ed.), *Advances in man-machine system research* (pp. 47-92). Greenwich, CT: JAI Press.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: models and methods* (pp. 138-147). New Jersey: Ablex Publishing Corporation.

Klein, G. A. (1997a). An overview of naturalistic decision making applications. In C. E. Zsombok, & G. Klein (Eds.), *Naturalistic decision making* (pp. 49-59). Mahwah, NJ: Lawrence Erlbaum.

Klein, G. A. (1997b). The current status of the naturalistic decision making framework. In R. Flin, E. Salas, M. Strub, & L. Martin, (Eds.), *Decision-making under stress: Emerging themes and applications* (pp. 137-146). Aldershot: Ashgate.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. *Proceedings of the Human Factors Society 30th Annual Meeting, 1*, 576-580.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor. *Journal of Management*, 20, 403-437.

Klinger, D. (2000). Observing teams in natural settings: A helpful tool to aid in team observation and feedback. *Paper presented at 5th conference on naturalistic decision making*.

Krahenbuhl, G. S., Marett, J. R., & Reid, G. B. (1978). Task-specific simulator pretraining and in-flight stress on student pilots. *Aviation, Space, and Environmental Medicine*, 49, 1107-1110.

Kraiger, K., & Wenzel, L. (1997). Conceptual development and empirical evaluation of measures of shared mental models as indicators of team effectiveness. In M. T. Brannick, E. Salas, & C. Prince, (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 63-84). NJ: Lawrence Erlbaum.

Krane, V., Anderson, M. B., & Strean, W. B. (1997). Issues of qualitative research methods and presentation. *Journal of Sport and Exercise Psychology*, 19, 213-218.

Krustrup, P., & Bangsbo, J. (2001). Physiological demands of top-class soccer refereeing in relation to physical capacity: effect of intense intermittent exercise. *Journal of Sport Sciences*, 19, 881-891.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Langan-Fox, J., Code, S., & Langfield-Smith, K. (2000). Team mental models: Techniques, methods and analytic approaches. *Human Factors* 42, 242-271.

Loehr, J. E. (1993). *Toughness training for life*. NY: Dutton/Plume.

Lipshitz, R. (1993). Decision making as argument-driven action. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: models and methods* (pp. 172-181). New Jersey: Ablex Publishing Corporation.

Lipshitz, R., & Ben Shaul, O. (1997). Schemata and mental models in recognition-primed decision making. In C. E. Zsombok, & G. A. Klein, (Eds.), *Naturalistic decision making* (pp. 293-303). New Jersey: Lawrence Erlbaum.

Lyle, J. (1999). Coaching and management of performance systems. In N. Cross, & J. Lyle, (Eds), *The coaching process* (pp. 233-254). Oxford, England: Butterworth-Heinemann.

MacMahon, C., & Ste-Marie, D. M. (1999). *Decision-making in rugby officials*. Paper presented at the Canadian Society for Psychomotor Learning and Sport Psychology, Edmonton: Alberta.

Marshall, C., & Rossman, G. B. (1989). *Designing qualitative research*. Newbury Park, CA: Sage.

Martin, J. A., Tolfrey, K., Smith, N. C., & Jones, A. M. (2002). Development and validation of a referee-specific fitness test. *Proceedings of Commonwealth Games/BASES Conference, Manchester*, p. 242.

Mascarenhas, D. R. D. (2004). *The psychology of refereeing: Rugby football league coherence training results* (Tech. Rep.). University of Edinburgh, PESLS Department.

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (2002a). The art of reason versus the exactness of science in elite refereeing: Comments on Plessner and Betsch (2001). *Journal of Sport and Exercise Psychology*, 24, 328-333.

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (2002b). The 'four cornerstones' model, *Referee*, 27(4), 38.

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (2005), Elite refereeing performance: Developing a model for sport science support. *The Sport Psychologist*, 19, 364-379.

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (in press, 2005), Assessing the accuracy and coherence of decision making in rugby union officials, *Journal of Sport Behavior*, 28(3).

Mascarenhas, D. R. D., Collins, D., Mortimer, P., & Morris, R. L. (2005). Training accurate and coherent decision making in rugby union referees. *The Sport Psychologist*, 19, 131-147.

Mascarenhas, D. R. D., MacPherson, A., Ollis, S., & Collins, D. (2003, March) *The umpire strikes back: Keeping your elite referees legal*. Paper presented at British Psychological Society Annual Conference, Bournemouth, England.

McCaffrey, N., & Orlick, T. (1989). Mental factors related to excellence among top professional golfers. *International Journal of Sport Psychology*, 20, 256-278.

McCullagh, P. (1986). Model status as a determinant of observational learning and performance. *Journal of Sport Psychology*, 8, 319-331.

McCullagh, P. & Starek, J. (1999). The effect of self-modeling on the performance of beginning swimmers. *The Sport Psychologist* 13, 269-287.

McLennan, J., & Omodei, M. M. (1996). The role of prepriming in recognition-primed decision making. *Perceptual and Motor Skills*, 82, 1059-1069.

Means, B., Salas, E., Crandall, B., & Jacobs, T. O. (1993). Training decision makers for the real world. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 306-326). New Jersey: Ablex.

Melrose, A. (1998). More jaw/jaw, less war/war: Coach and referee communication. *RFU Journal*, (Autumn 1998), 12-13..

Melrose, A. (1999, March). Developing the x-factor of refereeing. *Paper presented at the south west national panel referees' meeting*. Taunton, UK.

Melrose, A. (2003). *RFU development referee course booklet*. Wolverhampton, England: RFU Referees Centre of Excellence.

Metcalf, K. (2001). *Consistency equals confidence*. The Yorkshire Federated Society of Referees Performance Conference (August 5<sup>th</sup>, 2001). AMS Training Centre, Strensall; from <http://www.sportnetwork.net/main/s23/st40797.htm> - retrieved April 6th 2005 - "Refereeing consistency is all we can hope for", QPR Supporters club.

Millgram, E., & Thagard, P. (1996). Deliberative coherence. *Synthese*, 108, 63-88.

Morrison, E. (2002, August). Refereeing performance and the role of the touch-judge. *Paper presented at the inaugural elite referee unit conference*. Huddersfield, UK.

Mortimer, P. W., & Collins, D. J. (1997, September). Coherence of decision-making in team sports. *Paper presented at the BASES annual conference*, York.

Nevill, A. M., Balmer, N. J., & Williams, M. (2002). The influence of crowd noise and experience upon refereeing decisions in football *Psychology of Sport and Exercise*, 3(4), 261-272.



- Nisbett, R. E., & DeCamp Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231-259.
- Noble, D., Boehm-Davis, D., & Grosz, C. G. (1986). *A schema-based model of information processing for situation assessment* (NTIS No. ADA163150). Vienna, VA: Engineering Research Associates.
- Noble, D., Boehm-Davis, D., & Grosz, C. G. (1987). *Rules schema and decision making*. Vienna, VA: Engineering Research Associates.
- Norman, D. A. (1983). Some observations on mental models. In D. Gentner, & A. L. Stevens. (Eds.), *Mental models* (pp 7-14). New Jersey: Lawrence Erlbaum.
- Omodei, M., McLennan, J., & Whitford, P. (1998). Using a head-mounted video camera and two-stage replay to enhance orienteering performance. *International Journal of Sport Psychology*, 29, 115-131.
- Omodei, M., Wearing, A., & McLennan, J. (1997). Head-mounted video recording: A methodology for studying naturalistic decision making. In, R. Flin, E. Salas, M. Strub. & L. Martins (Eds.), *Decision making under stress: Emerging themes and applications* (pp. 137-146). Aldershot, UK. Ashgate Publishing Ltd.
- Orasanu, J. M. (1990). Shared mental models & crew decision making. *Technical report for Princeton University*, 46, 19.
- Orasanu, J., & Connolly, T. (1993). The reinvention of decision-making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision-making in action: Models and methods* (pp. 3-20). Norwood, NJ: Ablex.
- Orlick, T. (2000). *In pursuit of excellence* (3<sup>rd</sup> ed.). Champaign, IL: Human Kinetics.

Orlick, T., & Partington, J. (1988). Mental links to excellence. *The Sport Psychologist*, 2, 105-130.

Oudejans, R. R. D., Verheijen, R., Bakker, F. C., Gerrits, J. C., Steinbrückner, M., & Beek, P. J. (2000). Errors in judging 'offside' in football. *Nature*, 404, 33.

Partington, J., & Orlick, T. (1987). The sport psychology consultant evaluation form. *The Sport Psychologist*, 1, 309-317.

Patton, M.Q. (2002). Qualitative research & evaluation methods (3<sup>rd</sup> ed.). Thousand Oaks, CA: Sage.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 534-552.

Plessner, H., & Betsch, T. (2001). Sequential effects in important referee decisions: The case of penalties in soccer. *Journal of Sport and Exercise Psychology*, 23, 254-259.

Poczwadowski, A., Sherman, C. P., & Henschen, K. P. (1998). A sport psychology service delivery heuristic: Building on theory and practice. *The Sport Psychologist*, 12, 191-207.

Poll, G. (2003). Working lives: The referee. *Business Life*, 51-54.

Rainey, D. W. (1995). Sources of stress among baseball and softball umpires. *Journal of Applied Sport Psychology*, 7, 1-10.

Rainey, D.W., & Duggan, P. (1998). Assaults on basketball referees: a statewide survey. *Journal of Sport Behavior*, 21(1), 113-120.

Rainey, D.W., & Hardy, L. (1997). Ratings of stress by rugby referees. *Perceptual and Motor Skills*, 84, 728-730.

Rainey, D. W., Hardy, L. (1999a). Assaults on rugby union referees: A three union survey. *Journal of Sport Behavior*, 22(1), 105-113.

Rainey, D. W., & Hardy, L. (1999b). Sources of stress, burnout and intention to terminate among rugby-union referees. *Journal of Sport Sciences*, 17, 797-806.

Rainey, D., & Larsen, J. D. (1988). Balls, strikes, and norms: rule violations and normative rules among baseball umpires. *Journal of Sport and Exercise Psychology*, 10, 75-80.

Rainey, D., Larsen, J. D., Stephenson A., & Olson, T. (1993). Normative rules among umpires: The "phantom tag" at second base. *Journal of Sport Behavior*, 16, 3, 147-155.

Rainey, D. W., & Winterich, D. (1995). Magnitude of stress reported by basketball referees. *Perceptual and Motor Skills*, 81, 1241-1242.

Ramsbottom, R., Brewer, R., & Williams, C. (1988). A progressive shuttle run test to estimate maximal oxygen uptake. *British Journal of Sports Medicine*, 22, 141-144.

Rasmussen, J. (1985). The role of hierarchical knowledge representations in decision making and system management. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-15(2), 234-243.

Riley, P. (1993). The winner within: A life plan for team players. New York: Putnam.

Roth, E. M. (1997). Analysis of decision making in nuclear power plant emergencies: An investigation of aided decision making. In C. E. Zsombok, & G. A. Klein, (Eds.), *Naturalistic decision making* (pp. 175-182). New Jersey: Lawrence Erlbaum.

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man and Cybernetics*, 22, 1296-1308.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: prospects and limits in the search for mental-models. *Psychological Bulletin*, 100, 349-363.

Rumelhart, D. D., & Ortony, A. (1977). The representations of knowledge in memory. In R. C. Anderson & R. J. Spiro (Eds.), *Schooling and the acquisition of knowledge* (pp. 99-135). Hillsdale, NJ: Lawrence Erlbaum.

Sabatini, D. (2002). The search for excellence. *Referee*, 24-28.

Salas, E., Cannon-Bowers, J. A., & Johnston, J. H. (1997). How can you turn a team of experts into an expert team?: Emerging training strategies. In C. E. Zsombok, & G. A. Klein, (Eds.), *Naturalistic decision making* (pp. 359-370). New Jersey: Lawrence Erlbaum.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors*, 27, 285-300.

Seel, N. M., Al-Diban, S., & Blumschein, P. (2000). Mental models and instructional planning. In M. Spector, & T. M. Anderson, (Eds.), *Integrated and holistic perspectives on learning, instruction and technology: Understanding complexity* (pp. 129-158). Dordrecht, NL: Kluwer Academic Publishers.

- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99-118.
- Simon, H. A., & Chase, W.G. (1973). Skill in chess. *American Scientist*, 61, 393-403.
- Sloan, T. (2004). Refs can't rate refs. *Referee*, 329, 58-61.
- Starkes, J. (2000). The road to expertise: Is practice the only determinant? *International Journal of Sport Psychology*, 31, 431-451.
- Starkes, J. L., & Allard, F. (1993). *Cognitive issues in motor expertise*. Amsterdam, Holland: Elsevier.
- Starkes, J. L., & Lindley, S. (1994). Can we hasten expertise by video simulations? *Quest*, 46, 211-222.
- Ste-Marie, D. (1999). Expert and novice differences in gymnastic judging: An information processing perspective. *Applied Cognitive Psychology*, 13, 269-281.
- Ste-Marie, D. (2003). Expertise in sport judges and referees. In J. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 169-189). Champaign, IL: Human Kinetics.
- Stern, J. (2002). Evaluating officiating performance. *Referee*, 27(9), 63.
- Stewart, M. J., & Ellery, P. J. (1996). Amount of psychological stress reported by high school volleyball officials. *Perceptual and Motor Skills*, 83, 337-338.
- Stokes, A. F., Kemper, K., & Kite, K. (1997). Aeronautical decision making, cue recognition, and expertise under time pressure. In C. E. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 183-196). Mahwah, NJ: Lawrence Erlbaum.

Stout, R. J. Cannon-Bowers, J. A., & Salas, E. (1996). The role of shared mental models in developing team situational awareness: Implications for training. *Training Research Journal*, 2, 85-116.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage, London.

Taylor, A. H., Daniel, J. V., Leith, L., & Burke, R. J. (1990). Perceived stress, psychological burnout and paths to turnover intentions among sport officials. *Journal of Applied Sport Psychology*, 2, 84-97.

Taylor, J. (1995). A conceptual model for integrating athletes' needs and sport demands in the development of competitive mental preparation strategies. *The Sport Psychologist*, 9, 339-357.

The Guardian. (2001). *Referee Gallagher relegated after Keane blunder*, p. 15.

The Observer. (2002). *Whistle while you work*, p. 17 (May 19).

Thomas, K.W., & Kilmann, R.H. (1974). *Thomas-Kilmann conflict mode instrument*. Tuxedo, NY: Xicom.

Time Magazine. (2001, August 25). *The Villains in Blue*, 13-14.

Tolcott, M. A., Marvin, F. F., & Bresnick, T. A. (1989). *The confirmation bias in military situation assessment*. Reston, VA: Decision Science Consortium.

Topp, W. (1999). Managing Conflict. In J. Grunska, (Ed), *Successful sports officiating* (pp. 61-84). Champaign, IL: Human Kinetics.

Torrance, E. P. (1953). The behaviour of small groups under the stress conditions of "survival." *American Sociological Review*, 19, 751-755.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453-458.

Webber, S. S., Chen, G., Payne, S. C., Marsh, S. M., & Zaccaro, S. J. (2000). Enhancing team mental model measurement with performance appraisal practices. *Organizational Research Methods*, 3(4, Oct 2000), 307-322.

Weinberg, R.S., & Richardson, P.A. (1990). *Psychology of officiating*. Champaign, IL: Leisure Press.

Williams, A. M., & Davids, K (1995). Declarative knowledge in sport: A by-product of experience or a characteristic of expertise. *Journal of Sport and Exercise Psychology*, 17, 259-273.

Williams, A. M. Davids, K., Burwitz, L., & Williams, J. G. (1994). Visual search strategies of experienced and inexperienced soccer players. *Research Quarterly for Exercise and Sport*, 65, 127-135.

Williams, A. M., Davids, K., & Williams, J. G. (1999). *Visual perception & action in sport* (1st ed.). London: E and FN Spon.

Williams, A. M., & Grant, A. (1999). Perceptual skills in soccer: Implications for talent identification to enhance coach-performer interactions. *Journal of Sports Sciences*, 18, 737-750.

Wilson, R. (Producer & Director) (2000, July 25). *Man in black*. Scotland: British Broadcasting Corporation.

Yates, J. F. (2001). "Outsider": Impressions of naturalistic decision making. In E. Salas, & G. Klein (Eds.), *Linking expertise and naturalistic decision making* (pp. 9-33). Mahwah, NJ: Lawrence Erlbaum.

Yukelson, D. (1997). Principles of effective team building interventions in sport: A direct services approach at Penn State University. *Journal of Applied Sport Psychology*, 9, 73-96.

Zakay, D., & Wooler, S. (1984). Time pressure, training, and decision effectiveness. *Ergonomics*, 27, 273-284.

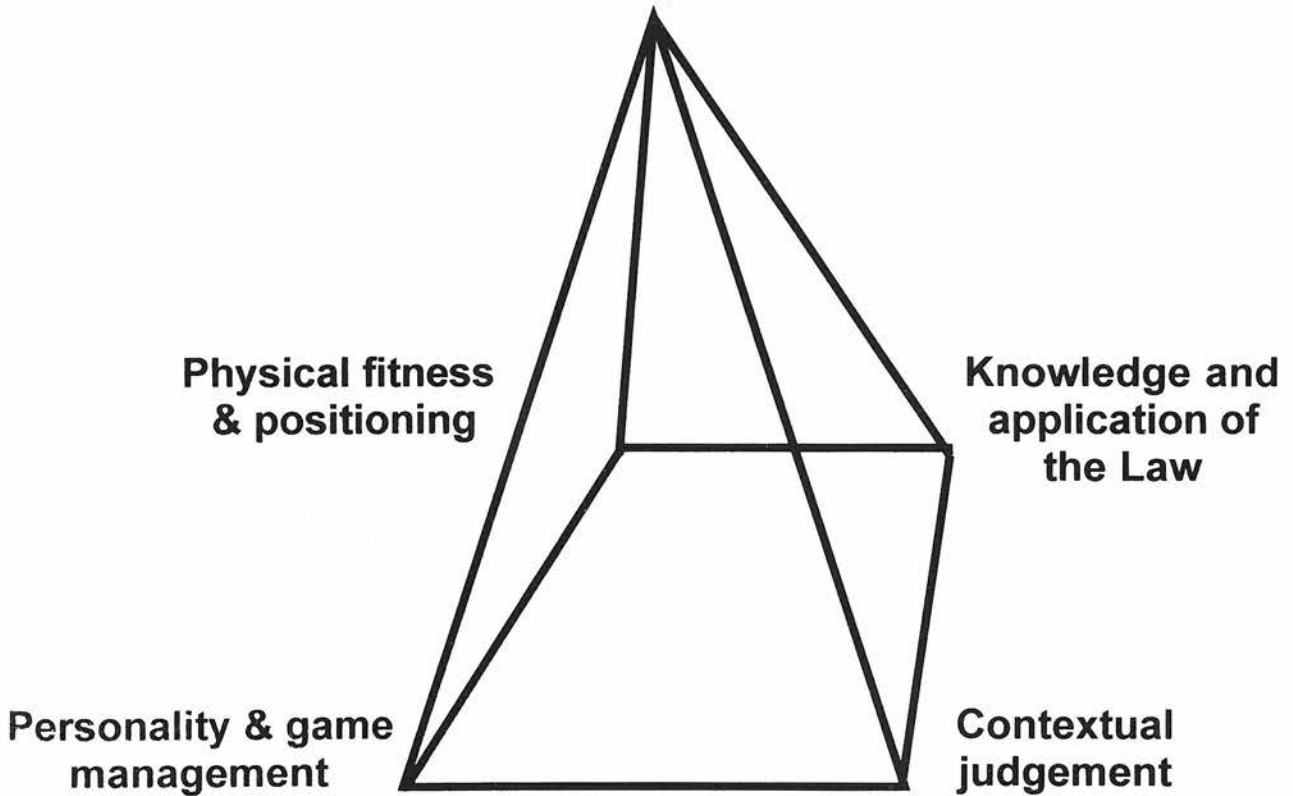
Zsombok, C. E. (1997). Naturalistic decision making: Where are we now? In C. E. Zsombok, & G. Klein (Eds.), *Naturalistic decision making* (pp. 3-16). Mahwah, NJ: Lawrence Erlbaum.



## Appendix One

1. Here is a graphic of the "Cornerstones Model of Refereeing". This is the structure we have used to study refereeing.

### The Psychological Characteristics of Excellence



**Knowledge and Application of the Law:** The underpinning knowledge of the law that allows referees to accurately interpret dynamic situations and penalise accordingly.

**Contextual Judgment:** An appreciation of the tenor of the game and the referee's ability to alter his or her style of refereeing to suit the particular nuances of the game.

**Personality and Game Management:** The verbal and non-verbal interpersonal skills that allow referees to communicate appropriately with the players.

**Physical Fitness, and Positioning:** The physical attributes that allow referees to "stay-up-with-play," see the game from an appropriate perspective and use the appropriate signals.

**Psychological Characteristics of Excellence:** The common elements of successful performance, including commitment, goal setting, imagery, planning, distraction control, responses to pressure situations and realistic performance evaluations.

- ♦ How appropriate do you think the model is for studying referee performance? Does it address all the key aspects of referee performance?

---

---

---

---

---

---

---

---

---

---

---

---

- ♦ How might you modify it?

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

**RESPONSE SHEET**

# CLIP 1

RESPONSE – Please tick the appropriate box.

- 1 ☐ **No action, play on/allow play to develop**
- 2 ☐ **Not enough information**
- 3 ☐ **Manage the situation**
- 4 ☐ **Advantage**
- 5 ☐ **Penalty awarded to Attacking team**
- 6 ☐ **Penalty awarded to Defending team**
- 7 ☐ **Free kick awarded to .....**
- 8 ☐ **Scrum awarded to Attacking team**
- 9 ☐ **Scrum awarded to Defending team**

Explain the reasons behind your decision.

.....

.....

.....

.....

.....

How was the quality of this clip?

.....

.....

.....

.....

How **confident** are you in your decision? (please circle the appropriate number)

not at all

completely

1

2

3

4

5

**COHERENCE STUDY 2 – RESPONSE SHEET**

Name.....

Please tick the appropriate box:                      ☐Rank (1-20)                      ☐Rank (21-47)                      ☐Rank (48-65)

**CLIP A1 – Leicester V Saracens**

RESPONSE – Please tick the appropriate box.

- 1 ☐ Play on
- 2 ☐ Penalty awarded to Leicester
- 3 ☐ Penalty awarded to Saracens
- 4 ☐ Scrum awarded to Leicester
- 5 ☐ Scrum awarded to Saracens
- 7 ☐ Other (explain).....

Explain the reasons behind your decision.

.....

.....

.....

.....

.....

.....

.....

What other information would you have liked, or how would you have preferred to view this tackle?.....

.....

.....

How **confident** are you in your decision? (please circle the appropriate number)

not at all				completely
1	2	3	4	5

## Appendix Four

Name ..... Current Ranking .....

### NORTHERN PANEL GROUP VIEWING LOG

The emboldened dates are the important ones. Please ensure that there is no more than a day either side of these dates when you actually view the video. I have suggested your regular night as Monday, but any night will do as long as stick to it. The model interpretations should not be looked at in the 3 days prior to the retest on the Final Response sheets.

**If you do not keep a stringent Viewing Log then the data that we receive will be of no use to us, and we would rather you completely stop the training all together.**

	IDEAL DATES	YOUR LOG OF ACTUAL DATES
<b>PRE TEST AT WORKSHOP</b>		<b>Sept 6<sup>th</sup></b>
B1-B5 Week 1 tackles First Response	<b>20<sup>th</sup></b>	
B1-B5 Model Interpretations video (week 1)	<b>20<sup>th</sup></b> (21 <sup>st</sup> , 22 <sup>nd</sup> , 23 <sup>rd</sup> )	
B1-B5 Final Responses (week 1 tackles)	<b>27<sup>th</sup></b>	
C1-C5 Week 2 tackles First Response	<b>27<sup>th</sup></b>	
C1-C5 Model Interpretations video (week 2)	<b>27<sup>th</sup></b> (28 <sup>th</sup> , 29 <sup>th</sup> , 30 <sup>th</sup> )	
C1-C5 Final Responses (week 2 tackles)	<b>Oct 4<sup>th</sup></b> before workshop	
D1-D5 Week 3 tackles First Response	<b>Oct 11<sup>th</sup></b>	
D1-D5 Model Interpretations video (week 3)	<b>11<sup>th</sup></b> (12 <sup>th</sup> , 13 <sup>th</sup> , 14 <sup>th</sup> )	
D1-D5 Final Responses (week 3 tackles)	<b>18<sup>th</sup></b>	
E1-E5 Week 4 tackles First Response	<b>18<sup>th</sup></b>	
E1-E5 Model Interpretations video (week 4)	<b>18<sup>th</sup></b> (19 <sup>th</sup> , 20 <sup>th</sup> , 21 <sup>st</sup> )	
E1-E5 Final Responses (week 4 tackles)	<b>25<sup>th</sup></b>	
F1-F5 Week 5 tackles First Response	<b>25<sup>th</sup></b>	
F1-F5 Model Interpretations video (week 5)	<b>25<sup>th</sup></b> (26 <sup>th</sup> , 27 <sup>th</sup> , 28 <sup>th</sup> )	
F1-F5 Final Responses (week 5 tackles)	<b>November 1<sup>st</sup></b>	
<b>POST TEST AT WORKSHOP</b>		<b>Nov 3<sup>rd</sup></b>

Write below any feedback on the quality/value/validity of the training package.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

To what extent are you likely to adopt this pattern of Decision Making during your upcoming games? *(please circle the appropriate answer)*

not at all				completely
1	2	3	4	5

### Instructions to Participants on Coherence Training

#### Introduction

Show 4-Cornerstones Model, and decision-making Model, explaining that law application is just the first part. For most of you we think we've brought you as close together as we need to. For others, maybe we need more training. Either way, your results on this test should indicate that for us. These are the other cornerstones and we realise that they are equally important and in fact perhaps more important at times. However we will look at those later. In the future we hope to be able to train CJ and PM factors.

1. Familiarise yourself with the response sheet (talk them through it – any questions)
2. Advise to make the decision in your head then fill out the response sheet by ticking the appropriate box.
3. Attacking team I have considered it to be the team last in possession before the tackle, however in all cases it is clearly marked on your sheet – please go by that.
4. Note if your decision is to play advantage you also need to indicate the decision that you are playing advantage from.
5. Note confidence level, & difficulty (out of 10)
6. Explain Reasoning – eg..... Do not describe what is on the Master Response list - ie it does exactly what it says on the tin. Indicate why. Don't just say "well...he held on too long" – what was it about the incident that made it too long? ABBREVIATE YOUR ANSWER!!
7. First there will be 8-10 tackles to orientate yourself into the "mode" of refereeing.
8. Then there will follow 2 practice clips.
9. Title on screen will indicate when you need to make your decision.
10. **Your decision must be instinctive and immediate (if you don't have an instinctive and immediate decision, then you should tick play on or did not see.**
11. The tackles are difficult. You must make your best attempt. There will not always be sufficient information. Can you always see the ball when you're refereeing the tackle – no. Equally you will not always be able to see the ball in this set of tackles. Also, the FTR's have adjudged all these tackles as realistic game scenarios that represent the sort of decision you might be confronted with on the park.
12. Finally, all your responses will remain **confidential** – the only publication of results will be made by group results.

## Appendix Six

Write below any feedback on the quality/value/validity of the training package.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

To what extent are you likely to adopt this pattern of Decision Making during your upcoming games? *(please circle the appropriate answer)*

not at all				completely
1	2	3	4	5



# Appendix Seven

## Chapter 6 Focus Group Schedule

### FACILITATOR'S NOTES

1. Today's aim is to investigate **factors that affect your decision making** as a referee.
2. **Video, discuss, weight and pilot-test** their impact.
3. **Scientific study** – so my role is to gain an understanding of your perceptions, without 'leading you'.
4. **Video** of Clips **15, 16, 17**, different individual decisions, 16, 17, **timing, Referee B** other factors etc.
5. **Individually** list Contextual Factors (CFs)
6. **Master** list of CFs
7. **Chart** CFs into graphic to determine in-game and pre-post game.
8. **Collapse** list into **in-game** factors, with **priority** column

**BREAK** – type up flipchart onto individual sheets

1. **Individually weight** the list of in-game CFs on importance & write the mean group scores on the master list.
2. **Identify top 3 for each.**
3. **Debrief**

## Chapter 6 Focus Group Instructions to Participants

1. Give them several **warm up tackles** of level 4 especially. Show them video **clips 15, 16, & 17.**

Ref 1 F1, J, J  
Ref 2 F1, J, B  
Ref 3 F1, J, B  
Ref 4 R, J, J  
Ref 5 F1, J, B

1. Ask all to make a decision on each of the 3 clips.
2. Are your decisions the same as the last time? If not, why not? What has changed? Everything's absolutely identical isn't it?
3. Clip 17, why have some played on and others penalised the ball carrier? Look at it again. Can you see Referee A's decision now? If you have chosen to play on, under what circumstances might you make the other decision? Are they both plausible? [focus on the facts only] (If never...under what circumstances might you give another decision...if still never – okay next clip).
4. Raise arm when you've made your decision for clip **16 & 17**. Then indicate your decision. Was the timing of your decisions the same? Why not? Would your timing always be the same?
5. Show them Referee B's clip and raise arm when whistling. What decision do you think Referee B gave? Actually, he gave a penalty for stamping. Why do you think he did this? Where else might he get information to reverse his decision?
6. Do these types of differences occur/exist when you're out on the park? What other differences might there be out there on the park?
7. If you're making a decision for a stropky player would you do it in the same manner as the 'nice guy' who's showing you respect?

2. Individually List CF's on paper.	<b>What might occur around those tackles (either before or afterwards) to alter your decisions? How might you feel it necessary to make a different decision, or make a decision in a certain manner? Might you referee this situation differently in different game or for different individuals? If so, why and in what ways would you modify your decision-making behaviour?</b>
3. Make Master list on Flipchart.	List all without judgement.
4. Chart CF's into graphic.	<ol style="list-style-type: none"> <li>1. How might you categorise these? Can you arrange them by time. When do these events take place?</li> <li>2. List as pre-game, in-game, post-game (and in clip or other if they arise – eg TJ intervention)</li> <li>3. Use ABC/abc model if necessary</li> </ol>
5. Collapse list into Master with <b>IMPORTANCE</b> column	1. Re-list onto flipchart with <b>IMPORTANCE</b> column. Ask for eg, is this similar to this? Is this related to this? Does this fit into another factor?
<b>BREAK</b>	1. Type up master list onto PC and printout a copy each
6. Individually weight the CF's	<ol style="list-style-type: none"> <li>1. How important is that factor, how much might it affect your decision?</li> <li>2. How frequently does that occur?</li> <li>3. Input all their scores onto my laptop &amp; transfer priorities to master list.</li> </ol>
7. Identify top 3	<ol style="list-style-type: none"> <li>1. Check for top 3 GROUP, top 3 INDIVIDUAL</li> <li>2. I am interested in the group's (average) response, because ultimately we want the group to cohere. This should be your target, so you need to look at <b>where you differ</b>, and most importantly <b>why you differ</b>...</li> </ol>
8. Show same video clips with different antecedents.	<ol style="list-style-type: none"> <li>1. This final bit is exploratory (PILOT) to see how we might test the impact of CFs upon CJ.</li> <li>2. 1 clip each (first decision, then decide again after model interpretation, then add in one CF.</li> <li>3. Look at the <b>timing</b>, <b>selling</b> and <b>reasons</b> for changes?</li> </ol>
9. Debrief	<ol style="list-style-type: none"> <li>1. Awareness of CFs</li> <li>2. Awareness of personality factors</li> <li>3. Understanding of how CF's might affect your decisions</li> <li>4. Ideally we want coherence of pure law and CJ but recognise that personality will affect both of these.</li> <li>5. There are a number of solutions to making decisions in games – The aim is to send out a coherent message (ie you are fathomable) and make decisions that are appropriate for your personality/style.</li> </ol>

## Chapter 6 Focus Group Response Sheets

### List of Contextual Factors

What might occur around these tackles (either before or afterwards) to alter your decisions? Might you referee this situation differently in different game or for different individuals? If so, why and in what ways would you modify your decision-making or behaviour?

#### Clip 15 – Initial Decision..... Decision after model interpretation.....

Decision	
Reason	
Did your timing change? How?	
What might you say differently?	
Reasons for any changes?	

#### Clip 16 – Initial Decision..... Decision after model interpretation.....

Decision	
Reason	
Did your timing change? How?	
What might you say differently?	
Reasons for any changes?	

#### Clip 17 – Initial Decision..... Decision after model interpretation.....

Decision	
Reason	
Did your timing change? How?	
What might you say differently?	
Reasons for any changes?	

## Appendix Eight

Tackle	GREEN	AMBER			RED			Y-CARD	O-Card	R-CARD
	1	2	3	4	5	6	7	8	9	10
Counter	How illegal was that?		Decision?							
	How hard was your response? (eg verbal, non-verbal, whistle etc)				NONE	SOFT		MEDIUM	HARD	
Why was this your level of response?										

Tackle	GREEN	AMBER			RED			Y-CARD	O-Card	R-CARD
	1	2	3	4	5	6	7	8	9	10
Counter	How illegal was that?		Decision?							
	How hard was your response? (eg verbal, non-verbal, whistle etc)				NONE	SOFT		MEDIUM	HARD	
Why was this your level of response?										

Tackle	GREEN	AMBER			RED			Y-CARD	O-Card	R-CARD
	1	2	3	4	5	6	7	8	9	10
Counter	How illegal was that?		Decision?							
	How hard was your response? (eg verbal, non-verbal, whistle etc)				NONE	SOFT		MEDIUM	HARD	
Why was this your level of response?										

## Appendix Nine

### *Peer Reviewed Publications*

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (2002). The art of reason versus the exactness of science in elite refereeing: Comments on Plessner and Betsch (2001). *Journal of Sport and Exercise Psychology*, 24, 328-333.

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (2005). Elite refereeing performance: Developing a model for sport science support. *The Sport Psychologist*, 19, 364-379.

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (in press, 2005). Assessing the accuracy and coherence of decision making in rugby union officials. *Journal of Sport Behavior*, 28(3).

Mascarenhas, D. R. D., Collins, D., Mortimer, P., & Morris, R. L. (2005). Training accurate and coherent decision making in rugby union referees. *The Sport Psychologist*, 19, 131-147.

### *Other Publications*

Mascarenhas, D. R. D., Mortimer, P. & Collins, D. (June, 1999). Helping the man in the middle: Coherence assessment report. *RFU Technical Report*.

Mascarenhas, D. R. D., Mortimer, P. & Collins, D. (June, 2000). Helping the man in the middle: Coherence training report. *RFU Technical Report*.

Mascarenhas, D. R. D., Mortimer, P. & Collins, D. (June, 2000). Helping the man in the middle: Video, CD-Rom and website feasibility report. *RFU Technical report*.

Mascarenhas, D. R. D. & Mortimer, P. (October, 2000). RFU Referee agility exercises (video).

Mascarenhas, D. R. D., Mortimer, P. & Collins, D. (January, 2001). Helping the man in the middle: Communication & conflict management report. *RFU Technical Report*.

Mascarenhas, D. R. D., Mortimer, P. & Collins, D. (January, 2001). Helping the man in the middle: A reassessment of coherence. *RFU Technical Report*.

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (2002). The 'four cornerstones' model. *Referee: Magazine for the National Association for Sports Officials*, 27(4), 38.

### *Academic Conference Presentations*

Mascarenhas, D. R. D., Collins, D., & Mortimer, P. (2000, December) *Shared mental-models: A naturalistic approach to training coherent decision-making in rugby-union referees*. Scottish Sport Science and Medicine Conference in association with Sportscotland, BASES & BASEM (presentation prize-winner, 2000, December 11).

Mascarenhas, D. R. D. (2000, April) *Assessing and training referee decision-making*. RFU Referees: Managing the next steps conference, Halifax, Yorks.

Mascarenhas, D. R. D., MacPherson, A., Ollis, S., & Collins, D. (2003, March) *The umpire strikes back: Keeping your elite referees legal*. Paper presented at British Psychological Society Annual Conference, Bournemouth.

### *RFU Workshops and Conference Presentations*

Mascarenhas, D. R. D. (1999, September). *Assessing referee decision-making: Results, and Performance profiling and POTS analysis*. RFU Referees regional meetings (x4) September 1999.

Mascarenhas, D. R. D. (1999, November). *Training referee decision-making*, and *An introduction to conflict management*. RFU Referees regional meetings (x4) November 1999.